



The Austrian Ludwig Wittgenstein Society
in cooperation with Central European University
and with the support of
the Federal Government of Lower Austria, section Science and Research,
and the municipalities of Kirchberg, Otterthal, and Trattenbach
is pleased to present the

45th International Wittgenstein Symposium
August 11–17, 2024

FACETS OF REALITY

CONTEMPORARY DEBATES

Organised by
Asya Passinsky (Vienna), Julio De Rizzo (Vienna)
& Benjamin Schnieder (Vienna)

BEITRÄGE
CONTRIBUTIONS

45. Internationales Wittgenstein Symposium
Kirchberg am Wechsel | 2024

45

WISSENSCHAFT · FORSCHUNG
NIEDERÖSTERREICH



FACETTEN DER WIRKLICHKEIT

FACETS OF REALITY

Beiträge der Österreichischen Ludwig Wittgenstein Gesellschaft

Contributions of the Austrian Ludwig Wittgenstein Society

Band XXX
Volume XXX

FACETTEN DER WIRKLICHKEIT – ZEITGENÖSSISCHE DEBATTEN

Beiträge des 45. Internationalen Wittgenstein Symposiums
11. – 17. August 2024
Kirchberg am Wechsel

Band XXX

Herausgeber

Yannic Kappes
Asya Passinsky
Julio De Rizzo
Benjamin Schnieder

in Zusammenarbeit mit Joseph Wang-Kathrein

Die Herausgeber danken Denis Kazankov und Martyna Meyer
für ihre große Hilfe bei der Vorbereitung dieses Bandes.

Produziert mit Unterstützung der Abteilung
Wissenschaft und Forschung (K3)
des Amtes der NÖ Landesregierung

WISSENSCHAFT • FORSCHUNG
NIEDERÖSTERREICH 

Kirchberg am Wechsel, 2024

Österreichische Ludwig Wittgenstein Gesellschaft

FACETS OF REALITY – CONTEMPORARY DEBATES

Contributions of the 45th International Wittgenstein Symposium
August 11 – 17, 2024
Kirchberg am Wechsel

Volume XXX

Editors

Yannic Kappes
Asya Passinsky
Julio De Rizzo
Benjamin Schnieder

in cooperation with Joseph Wang-Kathrein

The editors would like to thank Denis Kazankov and Martyna Meyer
for their substantial help with the preparation of this volume.

Produced with the support of the
Department for Science and Research (K3)
of the Province of Lower Austria

WISSENSCHAFT • FORSCHUNG
NIEDERÖSTERREICH 

Kirchberg am Wechsel, 2024

Österreichische Ludwig Wittgenstein Gesellschaft

Verleger | Publisher

Österreichische Ludwig Wittgenstein Gesellschaft

Austrian Ludwig Wittgenstein Society

Markt 63, A-2880 Kirchberg am Wechsel
Österreich/Austria

www.alws.at

Vorstand der ÖLWG | Executive Committee of the ALWS

Christian Kanzian (Präsident der ÖLWG)
Elisabeth Nemeth (Vizepräsidentin der ÖLWG)
Peter Kampits (Vizepräsident der ÖLWG)
David Wagner (Geschäftsführer der ÖLWG)
Marian David
Elisabeth Ehrenhöfer
Christoph Limbeck-Lilienau
Josef Mitterer
Volker Munz
Alois Pichler
Klaus Puhl
Friedrich Stadler
Paul Weingartner

ISSN 1022-3398 | ISBN 978-3-9505512-2-8

Refereed Periodical

All Rights Reserved

Copyright 2024 by the Authors

Copyright will remain with the author, rights to use with the society. No part of the material may be reproduced or utilised in any form or by any means, electronically or mechanically, including photocopying, recording, informational storage, and retrieval systems without permission from the society.

Die Redaktion dankt allen Gutachtern für ihre äußerst hilfreichen Rezensionen.

The editors are grateful to all referees for their most helpful reviews.

Die Beiträge und das Abstracts-Heft wurden mit Unterstützung einer von Joseph Wang-Kathrein, Universität Innsbruck, erarbeiteten Datenbank erstellt. Kontakt: joseph.wang@uibk.ac.at

The *Contributions* and the booklet of *Abstracts* were compiled with the support of a database developed by Joseph Wang-Kathrein, University of Innsbruck. Contact: joseph.wang@uibk.ac.at

Inhalt / Contents

The Right Language for Real Metaphysics	11
Siantonas Alec	
Degrees of Reality	20
Damian Aleksiev	
Is There a Dispositional Ground for the Contingency of Origin?	31
Jonas Amar	
Being Queer	41
Matthew Andler	
How Hegel brought Brandom back to Wittgenstein	43
Pavel Arazim	
“Will the Sun Rise Tomorrow?” Wittgenstein on Scepticism and Hinge Propositions	52
Lorenzo Barba	
Inside and Outside the Boundaries: Probability in Wittgenstein’s <i>Tractatus</i>	62
Matteo Bizzarri	
Toward a Dissolution of the Color-Exclusion “Problem”	74
Edgar Boedeker	
Being Pleonastic	86
Paolo Bonardi	
Are Artificial Neurons Neurons?	98
Johannes Brinz	
Wittgenstein’s <i>Tractatus</i>: A Transcendental Account	108
Cecilia Brotzu	
An Argument for Modal Realism	116
David Builes	
Connecting Language-Games: Diamond on Truth, Realism and The World	118
Barnaby Burleigh	
Logic as Metaphysics in Wittgenstein	127
Bangrui Chen	
Sortal Terms and the Puzzle of Coincidence	137
Zhengguan Chen	
The Modal Argument for the Anchoring Model in Social Metaphysics	147
Samuele Chilovi	
Translating the <i>Tractatus</i> and Tractarian Ethics	149
Jeffrey Patrick Colgan	

Vaidya's Method of Variation in Imagination Revisited	160
Violeta Conde & Concha Martínez-Vidal	
Levels of Abstraction As Family Resemblances From the Classical to the Quantum Mechanical Representation of Reality	168
Niccolò Covoni & Silvano Zipoli Caiani	
A Bipartite Reading of the Private Language Argument of Philosophical Investigations	178
Haiqiang Dai	
Ontology Reconstructed: Another Way of Being Neo-Carnapian	185
Caterina Del Sordo	
The Geach/Black Wittgensteined	196
Dušan Dožudić	
Luck, Disappearing Agents and Physicalism	208
Yaren Duvarci	
How Neo-Meinongians Take Ontological Questions About Fictional Entities Easy	216
Manuel Englert	
Imaginary Anthropologies. On Wittgenstein's Last Writings and Epistemic Relativism	224
Claudio Fabbroni	
Alexithymia and Neurodivergence: A Wittgensteinian Deconstruction	234
Mélissa Fox-Muraton	
Organic and Artificial Intelligence: Wittgenstein's Cultural Pessimism and Modern AI	243
Craig Frayne	
Are Social Constructs Fictions? Odd Terminology in Harari's <i>Sapiens</i>	251
Martin F. Fricke	
On Raising My Arm Automatically and Similar Actions: Implications of Automaticity for Free Agency	256
Christina Fritz	
The Regress of Necessity and Its Structure	266
Yaokun Fu	
Two Concepts of Possibility in <i>Tractatus Logico-Philosophicus</i>	275
Cloris C. Gao	
Translating Wittgenstein's <i>Tractatus</i> Into RDF	283
Jakub Gomułka	
Fundamentality, Existence, Totality: On Three Notions of Reality and the Landscape of Metaphysics	292
Dustin Goßens	

Quantifiers, Grounds, and Ontological Commitment	301
Martin Grajner	
Grounding without Essence	303
Zhiwei Gu	
Is Creeping Minimalism a Threat to the Realism-Expressivism Distinction?	310
Mohammadjavad Hajjalikhani	
Holistic Understanding as the Primary Epistemic Aim of Metaphysics	318
Jani Hakkarainen	
A Pursuit of Wittgenstein’s Usage of “Chimeras” in <i>Philosophical Investigations</i> §94	327
Shogo Hashimoto	
Does Grounding Make Sense? A Humanistic Approach to Ordered Ontology	337
Iain Haslam	
Towards a Wittgensteinian Hinge Ethics: A Hinge-Epistemological Approach to Absolute Value in Wittgenstein	348
Jackson Hawkins	
The World and How We Know It: Wittgenstein’s Realism and Metaphysics	357
David Hommen	
<i>Tractatus</i> and “Radical Translation” Superimposed	367
Herbert Hrachovec	
Priority Monism and the Possibility of Gunk	378
Alina Sophie Jacobs	
From Observable Regularities to Rule-Governed Practices: Exploring the Normative Facet of Social Reality	388
Amadeusz Just	
Quasi-objects	397
Christian Kanzian	
Shared Responsibility Does Not Follow from Shared Agency	407
Pelin Kasar	
The Exportability of Embedded Social Kinds	415
Denis Kazankov	
Dummett and the Reality of Pain	425
Gilad Kenan	
Possibilism and the “Wrong Kind of Object” Problem	435
Tingjiang Kuang	
A Heideggerian Reading of the <i>Tractatus</i>	445
David Lindeman	

A Private Language Argument against Phenomenal Concepts	456
Yaojun Lu	
Non-representationalism and Anti-Metaphysical Arguments	465
Shiman Luo	
Exploring the Empty Sign: Meillassoux’s Speculative Materialism and Wittgenstein's <i>Tractatus</i>	472
Jakub Mácha	
On Granularity Matching and Scale Mapping: A Novel Challenge for Multiple Realization	482
Matteo Mattersberger	
Kripke and Wittgenstein on Essence, Necessity, and the A Priori	491
Jacob McDowell	
Against the “Correspondence” Intuition	500
Susanna Melkonian-Altshuler	
Does the Pauli Exclusion Principle Support Modalism or Essentialism?	502
Luke Miller	
LLM Collapse As Rule-Following Failure	510
Julia Minarik & Patrick Fraser	
Coreference and Cofundamentality	521
Yang Minghui	
What Wittgensteinian GPTs Can’t Do	530
Filippo Mosca & Jakub Gomułka	
Vices and Virtues of Digital Realities: Exploring the Personal And Social Potential of Online Interactions	541
Carmen Mossner & Sven Walter	
Why Structuralism Cannot Defeat Scepticism	551
Alex Murphy	
A Wittgensteinian Perspective on Realism/Anti-Realism Debates	563
Ragna Talea Oeynhausen	
Reichenbach’s Relativised A Priori	572
Luca Oliva	
Real Responses vs. Judgments	584
C. Naomi Osorio-Kupferblum	
Vergleichende Analyse der <i>Lebensform</i> und <i>Lebenswelt</i> bei Wittgenstein und Habermas	593
Christian Pflügl	

Wittgenstein über Gesichtsraum und Grammatik	604
Maximilian Philipps	
Non-naturalist Moral Causation	614
Audrey Powers	
Der <i>Tractatus Logico-Philosophicus</i> und die Rede vom Verstehen	628
Moritz René Pretzsch	
Wittgenstein and Schopenhauer on Self and Will	637
Duncan Richter	
Metaphysics qua Metaphysics	644
Amedeo Robiolio	
A Puzzle About Grounding and Specification	651
Pauliina Rumm	
How to Avoid Metaphysical Magic – From Social Ontology to Mathematics	660
Jenni Rytilä	
On the Nexus of Rule-Following, Learning, and Accountability in LLMs	669
Mirco Sambrotta	
Die Papierkrone am Schachkönig – zur Metaphorik von Wittgensteins Sprachkritik	678
Alfred Schmidt	
Wittgensteins logischer Atomismus und Sokrates' Traum	687
Jan Schnalke	
Overcoming Modal Skepticism via Conceptual Engineering	697
Krzysztof Sękowski	
Regularities and Social Practices: Reconsidering Wittgenstein's Rule-Following Argument	708
Kanit (Mitinunwong) Sirichan	
Signalling Systems: Logic as a Conventional Facet of Reality	719
Emelia Stanley	
Betrieb Wittgenstein am Ende doch nur Metaphysik?	729
Matthias Statzkowski	
Somewhere Between Eden and the Matrix: Variations on Themes from David Chalmers	739
Twan Stiegel	
The Part of the <i>Tractatus</i> That Anscombe Thought Is “Most Obviously Wrong”	748
Inbal Tamir	
There Is Still Work for the Fundamental	757
Tarun Thapar	

Agency in Virtual Environments	
Sabine Thuermel	766
Modal Realism and the PSR	
Tarik Tijanovic	772
Engineering Concepts: Of Idling and Hot-Running Wheels	
Jasmin Trächtler	779
Three (Argumentative) Cheers for Historiographic Realism	
Aviezer Tucker	789
Not All Kids Get Medals	
Bram Vaassen	799
A Response to Sider on Bare Particulars	
Alice van't Hoff	810
Why No Digital File Is Unique	
Joseph Wang-Kathrein & Ulrich Lobis	818
A Case for Gender Abolition	
Benjamin Weislogel	825
Can Wittgensteinian Use of Language Explain (Away) Intentionality and Mental Representation?	
Christian Helmut Wenzel	832
Overcoming McDowell's Dualism of First and Second Nature	
Yara Windmüller	841
Is There a Plausible Realist Theory of Fictional Characters?	
Andrew Wynn Owen	850
No Reduction of Slot-Filling to Grounding and Essence	
Li Xiao	859
Is Aboutness Commitment Possible?	
Peihong Xie	868
Contingentism, Grounding and the Reduction of Metaphysical Necessity to Essence	
Qichen Yan	878
The <i>Tractatus</i> and Satisfaction	
Kevin Yao	887
Coinciding Groups: A Higher-Level Pluralist Approach	
Yiwen Zhan	895
A Wittgensteinian Survey of Ethical Boundaries	
Muhammad Zulnoorain	905

The Right Language for Real Metaphysics

Siantonas Alec (Sonipat, India)

Abstract

This paper considers the predicament of someone who is puzzled about a facet of reality: in this case, composition. What language would be best for working through their puzzlement? Four options are considered: Ontologese, a language stipulated to hew to the fundamental metaphysical facts; Ordinary English, the vehicle of everyday speech; Theorese, a maximally expansive language designed to accommodate reference to an unconstrained abundance of objects; and Logic, which is to say any well-defined and understood formal language. Methodological doubts are raised about each of the first three in turn. Ontologese is overly focused on simplicity, and risks ignoring much of our relevant evidence; Ordinary English is overly-focused on common sense, and risks discarding theoretical elegance; Theorese is overly-focused on explanatory power, and risks discarding theoretical simplicity. Logic, meanwhile, achieves an ideal balance between all of these theoretical virtues, inheriting the strengths and addressing the weaknesses of its rivals. Concerns that it risks changing the subject are answered.

Wanda is puzzled about some facet of reality: the way that bricks and mortar can sometimes make a house, roots and branches a tree. ‘What’, as Wanda expresses her puzzlement, ‘is going on here? What is composition?’ Wanda wants to work through her confusion towards understanding, but recognises that this topic is so puzzling that it requires a great deal of delicacy even in framing the questions, let alone finding answers. What are the right words? Is she so much as speaking the right language?

This paper will consider four different languages in which Wanda might choose to work out her puzzlement. First is Ontologese, a language stipulated to be maximally sensitive to underlying metaphysical reality. Then there is Ordinary English, supposed to be the language we all speak (at least, if we are native English speakers - others will have their own equivalents) outside of the specialised contexts set aside for discussion of such metaphysical questions. Another, more speculatively identified, is Theorese, a language stipulated to be maximally responsive to the demands of scientific enquiry. The fourth, an old answer which has made a recent resurgence with a new twist, is Logic: for these purposes, any well-defined and understood formal language.

We will review a basic case for each language, and identify a guiding methodological principle with which it is associated. More critically, we will identify for each a hostage to theoretical fortune. But Wanda is most

concerned with her initial puzzlement about composition, and isn't familiar with these technical details. Given her fairly naive, methodologically uncommitted starting point, in what language should she have most confidence as a tool for working through her perplexity? I will argue that Ontologese, Ordinary English, and Theorese are unsatisfying in various ways, and that Logic is left as the most promising candidate, combining the strengths and addressing the weaknesses of the others.

1. Ontologese

Wanda wants to understand composition itself, not just what people usually say and think about it. But the language she habitually uses has been shaped by the unreflective assumptions and practical concerns of a host of ordinary folk whom she can hardly trust to think for her. So perhaps she should stipulate her way to a new language, Ontologese, in which truth and falsity is determined entirely by the fundamental facts of metaphysics (Dorr 2005; Sider 2009, 2011). She need no longer be misled by peculiarities of ordinary usage that are really about convenience, rather than fundamental truth. She would be using an instrument designed for the singular purposes of metaphysical enquiry.

Wanda prefers theories to be simple, and so would like to end up with a simple account of composition. Ontologese looks set to deliver one (Sider 2013). For one, it is beholden to fundamental metaphysical reality, and Wanda suspects this preference for simple theories makes most sense given that fundamental reality is simple. Beyond this, the theories stated in Ontologese need take no account of what we would say in everyday speech, being solely a vehicle for fundamental theorising. Thus the complications arising from our ordinary judgements are excised, simplifying what remains.

Wanda recognises that Ontologese has a hostage to theoretical fortune: the assumption that there is some layer of fundamental metaphysical reality open to investigation. While she is intrigued by the claim, she is in no position to assess it now, and wants meanwhile to press on through her perplexities concerning composition. Nevertheless, she has independent misgivings about Ontologese. At what cost does its simplicity come? If Wanda ignores the judgements she would render in ordinary speech, is she not throwing away

her evidence? (Williamson 2007) There is much she takes herself to know which bears on questions of composition: that there came to be a house where there had only been bricks, that the ship sailed safely into port through the storm.

That this store of knowledge is accessed in a natural language shaped by practical concerns is beside the point: it remains part of her evidence, and in considering other, less abstruse, questions, she would consider it folly to ignore so much of it. Surely it is the worse folly here, where she needs as much guidance through her puzzlement as she can get? At least as much as she wants them to be simple, Wanda wants her theories to be constrained by the evidence. Polishing those theories to a fine sheen of simplicity in isolation from it is a futile exercise: unless she takes her account of her evidence, she is not properly attending to the facet of reality which initially puzzled her.

2. Ordinary English

Given these worries, Wanda might be interested in Ordinary English. Those advocating it as a language for metaphysics, such as Amie Thomasson (2014) and Eli Hirsch (2011), profoundly oppose Ontologese. Designing a language that is perfectly sensitive to fundamental metaphysical reality is a mistake, since there is no fundamental metaphysical reality. Wanda may be genuinely perplexed by composition, but the solution to such perplexity is not a hubristic attempt to limn the ultimate fabric of things from the armchair, but simply to pay proper attention to the ordinary use of words. Hence to settle supposedly metaphysical perplexity, we had better be resolute in our commitment to Ordinary English.

If Wanda has spent enough time wrestling with questions of composition, they might well start to ring hollow for her, prompting her to ask whether they might be merely verbal. She might also be impressed by the thought that, despite her initial ambition to understand composition itself, the armchair is after all a much better vantage from which to survey language than fundamental reality. Beyond this Wanda may suspect that our ordinary ways of talking are just fine: that ‘There are mountains’ and ‘The ship sailed safely into port through the storm’ are plain truths as we speak them, and not responsibly abandoned in search of supposedly deeper truths.

Approaching metaphysical questions in Ordinary English brings with it a clear methodology: ‘Attend to the nuances of quotidian speech’ is already a methodological prescription. While it may be an imposition to speak of ‘theoretical virtues’ in this connection, we can identify a concern which plays the central role that simplicity does for Ontologese: conformity with common sense. What the folk assert and believe is the great test of any metaphysical claim, and Wanda is at least sympathetic to this approach, having already faulted Ontologese for diverging too sharply from it.

Wanda again recognises a hostage to theoretical fortune: the claim that there are no deep metaphysical facts (Hawthorne 2009). Again, she sets this aside to focus on concerns about how well this language is likely to help her puzzlement about composition. Why, Wanda wonders, should we be so invested in ordinary English? While she wants to respect ordinary composition talk, she may well find it muddled. She seeks elegant answers to her questions, whether or not this preference should have a basis beyond personal aesthetics, and so rejects muddles. Artists like Shakespeare and Milton innovated freely on the ordinary speech of their day in pursuit of beauty: why should Wanda not do likewise? If departing from Ordinary English might prove the price of elegance, then Wanda will be wary about making any advance commitment not to pay.

3. Theorese

So how might Wanda achieve elegance by departing from Ordinary English? Well, some think that there are many more composite objects than English speakers typically acknowledge, objects with both asteroids and chins as parts. Proponents of Ordinary English often accept that, if such people were foolhardy to form their own linguistic communities, their utterances of ‘There are objects with both asteroids and chins as parts’ would be true. But Wanda finds truth-conditional compositional semantics elegant, and so, if she were to grant that some utterances of ‘There are objects with both asteroids and chins as parts’ are true, would want to appeal to objects satisfying the predicates ‘has asteroids as parts’ and ‘has chins as parts’ to explain this. In general, for any

more expansive language, acknowledging more objects than Ordinary English, in which others might utter truths, she would want to interpret that language within an equally expansive meta-language (Eklund 2009).

This position is related to Neo-Fregean views in the philosophy of mathematics (Hale and Wright 2009; Eklund 2009; Russell 2017). Hume's law tells us that the number of xes = the number of ys if and only if there are exactly as many xes as there are ys. This bridge principle tells us when sentences containing syntactically singular number-terms are true, and given that they are true, we know that those terms refer, without making any extra check on whether there exists a special class of objects, the numbers, to which they refer.

This helps our mathematical reasoning go smoothly, but we might extend such reasoning indefinitely, invoking covering principles for any putative class of objects. Take the outpees, rupee notes that are beyond the territory of the Republic of India, and which come into and go out of existence on the border. We cannot say in advance how talking about outpees will help us to describe and explain the world, so, Wanda might reason, we should use a more expansive version of English which refers to them just in case it might prove useful. But Wanda could have substituted any putative class of objects for outpees, so she might as well start trying to speak a maximally extended English to cover them all, which we can call Theorese.

The point of calling it Theorese is that this language has the resources to posit any class of entity that might help our descriptions and explanations go more smoothly. The theoretical virtue with which this language is associated is clearly explanatory power, and it is primarily because of the promise of easy explanations that Wanda might want to use it.

Wanda recognises that Theorese has a hostage to fortune quite similar to that of Ordinary English: assuming that the world imposes no constraints on what syntactically singular terms can refer, given the right bridge principle (Eklund 2007). But setting direct exploration of this aside, Wanda might feel similar misgivings as she did for Ontologese. She is not used to thinking within this maximally expansive language, and in preferring it over a more familiar alternative she might risk losing her evidence. In particular, she accepts sentences which seem to her to express evidence about what does not exist: for

instance, ‘There are no sakes’. Even though ‘He did it for her sake’ is a true sentence in which ‘sake’ is a syntactically singular term, Wanda is confident that ‘sake’ does not refer, since reality is without sakes. Shifting to Theorese, which is designed to enable inferences such as that from ‘He did it for her sake’ to ‘Sakes exist’, endangers such evidence of non-existence.

In contrast to Ontologese, however, Theorese is totally antithetical to considerations of simplicity. It posits all manner of thing, not ultimately on the basis that such posits are valuable, but on the basis that they are costless. But, as we have seen, Wanda wants a simple theory. If she were puzzled by the weather, she would not want just to posit any assorted meteorological phenomena, at least without establishing that such posits really are what the evidence best supports: why should she be so lax in positing mereological phenomena? Wanda might not have any settled account of why simplicity matters, but matter to her it does. As with her preference for elegance, it could rest on a purely aesthetic basis, but again it is not clear why this should make a difference. Wanda admires the beauty created by Shakespeare and Milton under the constraints of iambic pentameter. In seeking for a theory of composition, why should she not, in her own small way, follow them by pursuing beauty under greater constraint than Theorese allows?

4. Logic

A final option for Wanda to consider is Logic, understood here as some well-defined and understood formal language or other. The idea is that Wanda would aim to work out her theories of composition in this formal language, given a canonical interpretation of its logical vocabulary, and definitions for its non-logical vocabulary in the meta-language of English (Quine 1948; Williamson 2013). This meta-language will not be Ordinary English, but a flexible, theoretically-enriched English. Unlike either Theorese or Ontologese, however, this meta-language will differ only in small simple ways from the language in which Wanda first posed her metaphysical perplexities, as well as the language which we use to interpret formal languages in contexts where no one is explicitly trying to build metaphysical theories.

Like Ordinary English, Logic is something that Wanda can understand well. Crucially, if she uses her meta-language deftly, she will be able to move

smoothly between sentences of English, in which she takes most of her evidence to be stored, and the formal language whose sentences she seeks to evaluate. Like Ontologese, though, this is an expressly theoretical language, insulated from practical concerns. The sentences which Wanda will evaluate are not the very ones she is habituated to affirming or denying, and while she can translate to and from English, identifying the best translation is a matter of fine judgement. Like Theorese, this language is a flexible one with great expressive power. Wanda could introduce and define new predicates, add modal operators to a non-modal language, and higher-order quantifiers to a first-order language. Should she be so inclined, she may, for all we have said, adopt a well-defined and understood non-classical language. This language, however, would still be constrained in ways that Theorese is not. For one, the consequences of any given sentences are easily identified; for another, though we can define the predicate P as translating a world like ‘sake’ used as (part of) a singular term in true English sentences, there is no guarantee that $\exists xPx$ is true.

It is worth pausing to consider this last point more closely. The idea is that Logic allows for the articulation of ‘deep metaphysical facts’, and in particular that it is, in some sense, ‘answerable to the world’ for its existence claims: like Ontologese, but unlike Ordinary English or Theorese. Why think this? Surely what is true in a formal language depends on the model used to interpret that language: boringly, $\exists xPx$ will be presumably true given that P is assigned the set of planets as its extension; more interestingly, it will be true even if it is assigned the set of sakes, given that there are sakes in the domain of discourse (see Chalmers 2009).

This is not the only way to think about the matter, though, and not obviously the best. I alluded vaguely above to the ‘canonical interpretation’ of logical vocabulary: in a crucial case, the quantifiers are interpreted as ranging unrestrictedly, over absolutely everything. Were sakes there to be quantified over, then $\exists xPx$ would be true, on our interpretation of the predicate P; but they are not, so it is not. Thus Logic is answerable to the world, particularly in respect of its quantifiers (Williamson 2013).

This, Wanda might think, is a hostage to theoretical fortune, but each of the languages we have considered so far has had such a hostage. Setting it, like the

others, aside, a worry of the sort that Wanda has had before is that moving to Logic is changing the subject. The questions she had initially asked were in English: either she would be left asking a new question entirely, or there is a mismatch between the logical answer and the original English question (Hofweber 2022).

Such worries, however, need not give Wanda too much pause. The original question simply is not that important to her: indeed, it was simply ‘What is going on here?’, and only incidentally happened to express her non-linguistic state of puzzlement at a particular facet of reality (compare *topics* in Capellen 2018). Part of which she wants is to aim for better questions. Of course, she may be mistaken about which questions are better, but there is little reason to expect that questions framed in Logic should be worse. Indeed, we have seen reasons to expect that they will be better: it is precise, powerful, its formal properties and relationship to English have been the subject of extensive study, and so on.

All told, Wanda should focus her efforts on Logic. She sees value in multiple theoretical virtues: simplicity, explanatory power, fit with common sense. Unlike the other options, dominated by a single virtue, Logic allows a balance between all three. Moreover, when comparing alternatives, a theory’s trying to hit three targets and succeeding tolerably well in all is a more interesting result than a theory’s trying to hit only one and succeeding perfectly. So insofar as Wanda may be puzzled about which among competing theories of composition to choose, Logic is a powerful tool for discrimination. seeking not only a good theory, but for ways to discriminate between theories, it is to Logic we should turn. For all that, the substantive commitments behind Logic may be wrong, and those of other languages may be right. But as a matter of which language’s hostage to fortune Wanda should hope to see redeemed, the answer is Logic.

References

Capellen, Hermann (2018) *Fixing Language*, Oxford: Oxford University Press.

Chalmers, David (2009) "Ontological Anti-Realism" in: David Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press.

Dorr, Cian (2005) "What We Disagree About When We Disagree About Ontology" in: Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics*, Oxford: Oxford University Press.

Eklund, Matti (2007) "Bad Company and Neo-Fregean Ontology", *Synthese* 170 (3), 393-414.

Eklund, Matti (2009) "Carnap and Ontological Pluralism" in: David Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press.

Hale, Bob & Wright, Crispin (2009) "The Metaontology of Abstraction" in: David Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press.

Hawthorne, John (2009) "Superficialism in Ontology" in: David Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press.

Hirsch, Eli (2011) *Quantifier Variance and Realism*, Oxford: Oxford University Press.

Hofweber, Thomas (2022) "The Case Against Higher-Order Metaphysics", *Metaphysics*, 5(1), 29–50.

Quine, W.V.O. (1948) "On What There Is", *The Review of Metaphysics* 2 (5): 21-38.

Russell, Jeffrey Sanford (2017) "Composition as Abstraction", *Journal of Philosophy* 114 (9), 453-470.

Sider, Theodore (2009) "Ontological Realism" in: David Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press.

Sider, Theodore (2011) *Writing the Book of the World*, Oxford: Oxford University Press.

Sider, Theodore (2013) "Against Parthood", *Oxford Studies in Metaphysics*, 8, 237–293.

Thomasson, Amie (2014) *Ontology Made Easy*, Oxford: Oxford University Press.

Williamson, Timothy (2007) *The Philosophy of Philosophy*, Malden: Blackwell.

Williamson, Timothy (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.

Acknowledgement

Thanks to Arya Amritansu and Kit Wilson for discussion.

Degrees of Reality

Damian Aleksiev (Vienna, Austria)

Abstract

This essay outlines a hierarchical framework of Reality that allows for degrees of Reality. I use Reality (with a capital “R”) to designate reality in a primitive, metaphysical sense. Reality, grounding, and essence are the key elements of the framework presented here. I assume that Reality must have a fundamental level and all fundamental phenomena must be Real. Moreover, I postulate that everything non-fundamental is ultimately grounded in the fundamental Real. But what about the Reality of the non-fundamental? I argue that it is possible for non-fundamental phenomena to be Real, Unreal, or Semi-Real. The framework developed here accommodates these possibilities and illuminates them using the notion of essence. I argue that the essential nature of a phenomenon determines its degree of Reality. The framework does not assume that Reality must have degrees but only that it may have degrees. Its theoretical attractiveness consists in its ability to accommodate many diverse intuitions about grounding, help us better understand and classify theories about grounding, and illuminate Reality and its possible degrees.

1. The Hierarchy of Reality

Reality appears to have a hierarchical structure. The less fundamental facts seem to be grounded in the more fundamental facts. It is plausible that all non-fundamental facts are ultimately grounded in fundamental and perfectly Real facts. These privileged facts might concern physical entities like particles or spacetime, or minds, or God, or something else. Let the fundamentalia be whatever they may be. What interests me in this essay is what bearing does the grounding relation have on the *reality of the non-fundamental*. I will articulate a framework where some non-fundamental facts might be perfectly Real, others Semi-Real, and yet others perfectly Unreal. In other words: I will demonstrate how *Reality might come in degrees*.

By “*Reality*” (with a capital R), I mean reality in the metaphysical sense. Reality, thus understood, is likely a primitive concept, as Fine (2001) believes it to be. Still, Reality can be illuminated in a few ways. For instance, it seems evident that existence is a necessary condition for being Real. Moreover, it is plausible that the Real should exist independently of our ideas, attitudes, or perceptions towards it. If so, being Real also requires representational independence.

The preceding glosses of the Real are not meant to be exhaustive. My proposal will focus on a further gloss. I will attempt to illuminate Reality via

considerations of *essence*. My preferred way to understand “essence” is inspired by the idea of a real definition, a definition of *what something is* (Fine 1994; Lowe 2012). For example, the essence of *being human* is to *be a rational animal*. I assume that all phenomena have essences, although, as I will show, some essences might be impossible to express in words.

Following the current orthodoxy, I understand *grounding* as a relation between facts. I will assume that facts are worldly entities. Specifically, every fact is a non-mereological unity constituted by one or more particulars and their properties or relations (Armstrong 1997). Every grounding connection obtains between ground/s and a groundee. The ground/s are always more fundamental than the groundee. They metaphysically determine and explain their groundees. Grounding, as I understand it, is asymmetric, transitive, and irreflexive.

The view I will be articulating takes for granted that there is a *fundamental level*: that some facts are absolutely fundamental. Moreover, it requires that *all* fundamental facts are Real.

Reality and grounding generate a hierarchical structure: a *hierarchy of being*. Fundamental Reality grounds a hierarchy of nodes ordered by grounding relationships. The nodes in this hierarchy can be Real, Unreal, or somewhere between the perfectly Real and the perfectly Unreal. The essential nature of every phenomenon determines its degree of Reality. Moreover, as I will demonstrate, it also determines the position of both Unreal and Semi-Real phenomena within the grounding hierarchy.

2. The Unreal

Any conception of Reality with a hierarchical structure requires that the fundamental facts are perfectly Real. But what about the derivative facts? Are the derivative facts likewise Real? Before going any further, I must note that grounding, as standardly conceived, is *neutral* on this question. As Fine (2001: 27) acknowledges, groundees might be either Real or non-Real. Nevertheless, many philosophers—for example, Cameron (2010), deRosset (2017), and Sider

(2011)–have assumed that the derivative facts fully lack Reality. Fine himself maintains that: “In the absence of any reason to the contrary [...] we should assume that any given grounded proposition is unreal.” (2001: 27)

I will refer to groundees that fully lack Reality as *Unreal*. Grounding is neutral on the Reality of groundees. Thus, the Unreality of a groundee cannot result solely from its role as a groundee; instead, it must result from additional factors. In this section, I will show that one good way to understand the Unreal is by considering essence.

Since grounding is a relation between facts, my focus will be on the essences of facts. For any fact F, I will understand F’s essence to be the fact corresponding to the answer to the question “what it is to be the case that F?” (Correia & Skiles 2019). For illustration, consider [Socrates is human]. It is reasonable to assume that for it to be the case that Socrates is human is for it to be the case that Socrates is a rational animal. If so, the essence of [Socrates is human] is [Socrates is a rational animal].

Many friends of grounding are what I call *top-down essentialists*. Top-down essentialists believe that many grounding connections are *mediated by the essence of the groundee*. Top-down essentialists include Aleksiev (2022), Dasgupta (2014), Fine (2012), Goff (2017), Jago (2018), Rosen (2010), and others. Top-down essentialism assumes an essential connection between groundees and their grounds. As Fine (2012: 76) puts it, the groundee’s essence, in some way, “points” to its ground. I offer the following characterization of the Unreal inspired by top-down essentialism:

The Unreal: a fact F is Unreal only if

1. F is ultimately grounded in a Real ground, and
2. Every possible ground of F contains an aspect identical to F’s full essence.

Clause (1) in the above characterization follows from the basic assumptions of my framework. Thus, my focus will be on illuminating clause (2).

Clause (2) states that the essence of every Unreal fact is identical to an *aspect shared by all possible grounds* of that Unreal fact. First, by “aspect,” I have in mind the product of *abstraction*, what we get after we abstract away from

some detail in an entity. Aspects are contained in the entities they are abstracted from. Second, I mention *possible grounds* because many groundees are multiply realizable; they can be grounded in many different grounds. Moreover, groundees often belong to grounding chains; they are separated from the Real by many derivative facts serving as mediate grounds.

Put another way, clause (2) says that for any Unreal fact F, F's essence corresponds to a condition that any possible ground of F must satisfy. This condition can be seen as the set-membership condition for being a member of the set of F's possible grounds.

The essences of Unreal facts are identical to aspects of their Real and ultimate grounds. Thus, it is adequate to say that the essence of every Unreal fact *fully points* to or, better put, *fully drains away* into its Real ground. This is why the Unreal facts lack Reality. They are exhaustively explainable by their Real grounds and are *nothing over and above* their Real grounds. An example might help. Consider the following fact:

Table: t is a table.

Table is likely grounded in some fact—call it Φ —about the macroscopic structure of *t*'s legs, surface, and other parts. In turn, Φ might be grounded in some fact Ψ about *t*'s chemical structure. And lastly, Ψ might be ultimately grounded in a Real fact Ω about *t*'s physical structure. Φ , Ψ , and Ω are three among many possible grounds for *Table*. *Table* is multiply realizable. It could have been grounded in many different arrangements of fundamental physical entities, or molecules, or macroscopic parts.

Now, consider *Table*'s essence. *Table*'s essence is likely a fact about *t* having the structural features of *being a table*. I propose that Φ , Ψ , and Ω , as well as any other possible ground of *Table*, share an aspect identical to *Table*'s essence. In other words, *Table*'s essence can be abstracted away from Φ , or Ψ , or Ω . Thus, *Table* is Unreal.

The above example is overly simple and lacks detail. Nevertheless, it illustrates my proposal. Moreover, its logic can be generalized to many other groundees.

Facts about ordinary objects, and presumably, also about societies, geology, living organisms, and so on, seem to be at least in principle subject to an analogous analysis. That is evidence of their Unreality.

3. The Real

Essence can also help us get a firmer grip on the Real. I stated that the Real facts can be fundamental or derivative. The fundamental facts *must* be Real, while the derivative facts *may* be Real. In this section, I will offer an essentialist gloss of both the fundamental and the derivative Real. Consider the following characterization.

The Real: a fact F is Real only if

1. F's essence is inexpressible or trivial or concerns other fundamentalia, and
2. F's essence is not identical to any substantive aspect of F's ground (if F is derivative).

Clause (1) intends to capture the intuition that Real phenomena are special because they have special essential natures. Clause (2) states that if a Real fact is grounded, its essence will never be identical with any substantive aspect of its ground. I say "substantive" because, for any two phenomena, no matter how distinct they may be, there will certainly be at least some aspects they have in common. For example, they will both be facts, or they will both exist. The substantiveness requirement is meant to exclude these aspects as obstacles to the grounding of Real facts.

Simply put: for any Real fact F, F's essence will either point *inwards*, to itself, or point to Real facts of *equal rank* to F. In either case, even if F is derivative, F's essence will never point to nor drain into F's ground. Thus, it is reasonable to say that the derivative Real facts will always be *fully over and above* their grounds.

Plausible candidates for Real facts include facts about fundamental physical entities, consciousness, God, moral and aesthetic value, and mathematical entities. I will illustrate my proposal with two examples: a form of physicalism and a form of dualism.

Consider *physicalism*: the thesis that all fundamental facts are physical. According to one influential view, the physical facts concern dispositional properties (Bird 2007). Dispositional properties essentially *are what they do*. Their essences plausibly are exhausted by their stimulus and manifestation conditions (Bird 2007: 45). These conditions will concern other dispositional properties that are fundamental and causally related to the property whose essence we are considering. If so, the essences of the fundamental facts in this view will likely only point to other fundamental facts. Thus, they will be Real.

Consider the following form of *dualism*: the thesis that both physical and mental entities exist, yet the physical entities are fundamental and ground the mental. Let us moreover assume that the physical entities are dispositional properties while the mental are phenomenal properties. I have already shown that, if fundamental, facts about dispositional properties are Real. I will now show that if this version of dualism is true, facts about phenomenal properties are likewise Real.

It is plausible that the essential natures of phenomenal properties concern their phenomenal characters; they concern what these properties are qualitatively like (Goff 2017). For example, the essence of red is what red *looks like*, while the essence of pain is what pain *feels like*. Such essences appear to be ostensive and, thus, inexpressible. If so, the essences of phenomenal facts will likely not point beyond themselves. There will be no essential connection between them and their physical grounds (Aleksiev 2022). Thus, this form of dualism posits both Real physical facts and Real phenomenal facts.

4. The Semi-Real

Reality is often viewed in a binary way. According to the *standard* view, Reality has no shades; things are either Real or Unreal, and there is nothing in between. In this section, I will sketch an alternative to the standard view. I will show that it is both coherent and in line with our intuitions that Reality might

come in degrees. Some facts may be *Semi-Real*. Essence will once again help us; this time, to understand the Semi-Real.

The Semi-Real: a fact F is Semi-Real only if

1. F is ultimately grounded in a Real ground, and
2. Every possible ground of F contains an aspect identical to a constituent of F's essence but never F's full essence.

Clause (1) should be obvious. Clause (2) expresses the idea that the Semi-Real facts are part Real and part Unreal. Like the Unreal facts, their essences point to or drain into their grounds. However, they *do not do so fully*. Like the Real facts, their full essence is non-identical to an aspect in their grounds. This makes the Semi-Real facts *something over above* their grounds, but *never fully*. They are, in a sense, *semi-reducible*. Consider the following two facts as examples:

Set: {Socrates} exists.

Heavy-Table: *t* is a table.

I begin with *Set*. *Set* is a fact about the existence of the singleton set {Socrates}. Intuitively, facts about the existence of sets are grounded in facts about the existence of their members. If so, *Set* is grounded in the fact that Socrates exists.

It is reasonable to assume that *Set*'s essence concerns Socrates' existence (Fine 1994). But will that be *Set*'s full essence? According to a *deflated* or *lightweight* conception of sets, sets are nothing more than their members or aggregates of their members (Armstrong 1991, 1997). However, it is also possible that sets are something more than their members, that sets are entities in their own *sui generis* way. Call this second conception *inflated* or *heavyweight*.

Set's essence will concern Socrates' existence on both the lightweight and the heavyweight conception. Thus, at least a part of *Set*'s essence will be identical to an aspect of *Set*'s ground. On the lightweight conception, that will be all there is to *Set*'s essence. However, there will be *more* to *Set*'s essence on the heavyweight conception. On the heavyweight conception, *Set*'s essence will contain a part about {Socrates} being a set. If so, *being a set* will be a

constituent of *Set's* essence. This constituent will not be identical to an aspect in *Set's* possible grounds. If so, *Set* is Unreal on the lightweight conception; however, it is Semi-Real on the heavyweight conception.

Onto *Heavy-Table*. In §2, I argued that facts about tables are Unreal, and I used a fact called *Table* to illustrate this. On the surface, *Table* and *Heavy-Table* are identical. However, just as we can have a lightweight or heavyweight conception of sets, so we can have of composites. I assumed a lightweight conception of composites in §2. In contrast, *Heavy-Table* assumes a heavyweight conception of composites.

On both conceptions, the essence of *Heavy-Table* will mention a structure that any possible ground of the corresponding fact must have as an aspect. However, according to the heavyweight conception, there will be more to *Heavy-Table's* essence. *Heavy-Table's* essence will contain a part about *t* being a composite (in the heavyweight sense). If so, *being a composite* will be a constituent of *Heavy-Table's* essence. This constituent will not be identical to an aspect in *Heavy-Table's* possible grounds. If so, *Heavy-Table* is Semi-Real while its lightweight relative *Table* is Unreal.

In summary: facts about sets and composites will be Semi-Real if sets and composites are heavyweight entities. The essences of such facts will be part Real and part Unreal. This distinguishes them from plausibly Real entities such as minds, fundamental physical particles, numbers, God, etc. Moreover, it also makes them distinct from Unreal entities such as lightweight composites or lightweight sets.

5. A Framework

I used essence and grounding to shed light on the Real, the Unreal, and the Semi-Real. My goal in this paper was not to argue for a specific worldview of what is Real, Unreal, or Semi-Real. Instead, I aimed to offer a *framework* that can accommodate many worldviews. The nodes in this framework can be filled or left empty depending on our views about what exists, what grounding relationships hold, and what the essences of the entities in these relationships are. Moreover, my goal was not to be exhaustive. Some phenomena may not fall under one of the three categories outlined here.

The unique feature of this framework is that it can accommodate a Reality with *degrees*. I did not argue that Reality must come in degrees nor that actual Reality comes in degrees. My goal was to demonstrate that Reality *may* have degrees.

The framework I presented assumes that all groundees are ultimately grounded in the Real. Moreover, it entails three categories of grounding connections. A grounding connection can fall under one of the following categories based on the essence of the groundee it involves:

1. *Reductive*: the groundee is Unreal.
2. *Semi-Reductive*: the groundee is Semi-Real.
3. *Non-Reductive*: the groundee is Real.

Reductive and semi-reductive connections are mediated by the groundee's essence. The mediation is full in reductive connections while partial in semi-reductive connections. Even in semi-reductive connections, the groundee's essence limits the possible grounds. In contrast, non-reductive connections—i.e., connections where both grounding partners are Real—are not mediated by essence. Instead, they must be either a matter of brute necessity or be mediated by (grounding) laws.

The above categories can help us better understand existing grounding proposals. For instance, Dasgupta (2014), Fine (2001), Rosen (2010), and Sider (2011) advocate for or lean towards the Unreality of the derivative. When taken all the way, views like these result in worldviews where all grounding connections are reductive. In contrast, Schaffer's (2017, 2021) system is best interpreted as only posting semi-reductive or non-reductive grounding connections. His system does not allow for Unreal groundees. Instead, it appears to require that all groundees are Semi-Real or Real.

I did not defend any of the foundational principles of this framework in this short essay. Moreover, I did not provide thorough support for the framework's elements. More work certainly remains to be done. Nevertheless, despite its current rough state, I believe the framework is theoretically attractive and motivated. It can accommodate a diverse spectrum of intuitions about grounding. It can help us better understand, classify, and compare current theories of grounding. It clarifies the notions of reduction and lack thereof, as

well as the notions of nothing and something over and above. Lastly, it illuminates the Real, the Unreal and the possible degrees of Reality between the Real and the Unreal.

References

- Aleksiev, Damian (2022) "Lightweight and Heavyweight Anti-physicalism", *Synthese*, 200(112). <https://doi.org/10.1007/s11229-022-03506-3>
- Armstrong, David Malet (1991) "Classes are states of affairs", *Mind*, 100(2), 189–200. <https://doi.org/10.1093/mind/c.398.189>
- Armstrong, David Malet (1997) *A World of States of Affairs*, New York: Cambridge University Press.
- Bird, Alexander (2007) *Nature's Metaphysics*, New York: Oxford University Press.
- Cameron, Ross P. (2010) "How to have a radically minimal ontology", *Philosophical Studies*, 151(2), 249–264. <https://doi.org/10.1007/s11098-009-9442-2>
- Correia, Fabrice; Skiles, Alexander (2019) "Grounding, Essence, And Identity", *Philosophy and Phenomenological Research*, 98(3), 642–670. <https://doi.org/10.1111/phpr.12468>
- Dasgupta, Shamik (2014) "The Possibility of Physicalism", *Journal of Philosophy*, 111(9–10), 557–592. <https://doi.org/10.5840/jphil20141119/1037>
- deRosset, Louis (2017) "Grounding the Unreal", *Philosophy and Phenomenological Research*, 95(3), 535–563. <https://doi.org/10.1111/phpr.12437>
- Fine, Kit (1994) "Essence and Modality", *Philosophical Perspectives*, 8, 1–16. <https://doi.org/10.2307/2214160>
- Fine, Kit (2001) "The Question of Realism", *Philosophers' Imprint*, 1(1), 1–30.
- Fine, Kit (2012) "Guide to Ground", in: Fabrice Correia and Benjamin Schnieder (eds.) *Metaphysical Grounding*, New York: Cambridge University Press, 37–80.
- Goff, Philip (2017) *Consciousness and Fundamental Reality*, New York: Oxford University Press.
- Jago, Mark (2018) "From Nature to Grounding", in: Ricki Bliss and Graham Priest (eds.) *Reality and its Structure: Essays in Fundamentality*, New York: Oxford University Press, 199–216.
- Lowe, Edward Jonathan (2012) "What is the Source of Our Knowledge of Modal Truths?", *Mind*, 121(484), 919–50. <https://doi.org/10.1093/mind/fzs122>
- Rosen, Gideon (2010) "Metaphysical Dependence: Grounding and Reduction", in: Bob Hale and Aviv Hoffmann (eds.) *Modality: Metaphysics, Logic, and Epistemology*, New York: Oxford University Press, 109–36.

Schaffer, Jonathan (2017) "The Ground Between the Gaps", *Philosophers' Imprint*, 17(11), 1–26.

Schaffer, Jonathan (2021) "Ground Functionalism", in: Uriah Kriegel (ed.) *Oxford Studies in Philosophy of Mind* (Vol. 1), New York: Oxford University Press, 171–207.

Sider, Theodore (2011) *Writing the Book of the World*, New York: Oxford University Press.

Acknowledgement

Many thanks to Paolo Bonardi, Yannic Kappes, Xiao Li, Julio De Rizzo, Benjamin Schnieder, and Ricardo Alcocer Urueta for helpful discussions. This research was funded in whole by the Austrian Science Fund (FWF) 10.55776/ESP318.

Is There a Dispositional Ground for the Contingency of Origin?

Jonas Amar (Paris, France)

Abstract

Could our reality have had a different origin? This paper examines a central claim for the modal dispositionalists: the necessity of origin. According to this theory (especially in Vetter (2015) and Kimpton-Nye (2021)) nothing has a disposition such that the origin of the universe could have been different. However, combined with other plausible claims, Werner (2022) and Vetter and Busse (2022) have shown that this could lead to *necessary perfect masks* and a form of *necessitism*. I suggest a possible solution to ground the contingency of origin within the “classic” modal dispositionalist account. For this, I argue that the necessity of origin relies on implicit assumptions concerning the debate after Kripke (1980) and his claim of the essentiality of origin. After recalling the two major interpretations of this claim, the sufficiency principles and the branching model, I show that the modal dispositionalist almost only rely on the latter. I can then present an argument for the contingency of origin: first I explain why the branching model isn’t fit for the problem of necessary perfect masks, and, then, I argue that one could find better resources in the debates concerning the sufficiency principles. If I am right, it is especially the *tolerance problem* that one can apply to the origin of dispositions. I conclude that one could thus find an ordinary disposition that is such that its manifestation is compatible with a limited set of possible different origins. Modal dispositionalism seems then compatible with the contingency of origin, a claim that would contradict *necessitism*.

1. Introduction: modal dispositionalism and the threat of *necessitism*

According to modal dispositionalism, every possibility is grounded in dispositions. This theory is especially presented and supported in Borghini and Williams (2008), Jacobs (2010), Vetter (2015). The central claim is the biconditional: p is possible iff something x has a disposition to be such that p . The idea is to ground every *de dicto* possibility in a *de re* claim. This is why Vetter (2015) suggests extending considerably the notion of disposition: it is not restricted to the individual and actual disposition, but could be applied to *joint* dispositions, *past* dispositions and *iterated* dispositions. One may thus suggest the dispositional biconditional:

DB – It is possible that p iff some things xx have, had, or will have an iterated or noniterated disposition for it to be the case that p .

With this considerable extension of the notion of disposition, the modal dispositionalist seems to have all the means she needs to account for all possibilities and necessities. That being said, the theory seems to face a major difficulty when it has to account for non-actualized possibilities. It seems to be

naturally committed to a form of *necessitism* (i.e. the thesis that every truth is necessarily true) as well noted by Vetter and Busse (2022) and Werner (2022a). This appears as a result of the conjunction of **DB** with three very plausible claims for the modal dispositionalist:

Necessity of origin – No *xx* have, had, or will have an iterated or noniterated disposition for the first state of the universe *H* to be different.

Determinism – The initial state of the universe *H* and the set *L* of all the actual laws of nature given, it is logically possible to deduce all true propositions about the state of the universe at a given time.

Necessity of laws – No *xx* have, had, or will have an iterated or noniterated disposition for the set *L* of all the actual laws of nature to be different.

The conjunction of those four claims contradicts the non-actualized possibilities:

Non-actualised possibilities – There is at least a proposition *p* which is possible but actually false.

Suppose then that *p* is a possibility non-actualized at t_1 . So, one has some *xx* that are such that *p* but actually $\neg p$. However, given the **Necessity of origin** and **Necessity of laws**, one has $\Box(H \wedge L)$. Given then **Determinism**, one should obtain $\Box((H \wedge L) \rightarrow \neg p)$. By the K-axiom, $\Box(H \wedge L) \rightarrow \Box \neg p$. This directly contradicts the idea that *p* is possible. One must then conclude that the conjunction of the **Necessity of origin**, the **Necessity of laws** and **Determinism** entails that there is no non-actualized possibility, and that every truth about our world is necessarily true.

This is a serious problem for the modal dispositionalist who seems then to be committed to *necessary perfect masks*. Some dispositions, such as the disposition of a glass to break, but which will never manifest itself, will *necessarily* never manifest itself. This limits the explanatory power of the theory: how can a disposition that cannot manifest itself still be considered a disposition for a certain manifestation? Worse still, and even more threatening to the theory, one has a disposition for *p* without the possibility that *p*.

Werner (2022a) suggests that the modal dispositionalist should extend his theory to include all the possibilities that are logically entailed by the manifestation of a certain disposition. If the glass broke, then it would be possible that $\neg(H \wedge L)$, even if there is no disposition that is such that $\neg(H \wedge L)$. Dispositions are then sufficient for grounding a possibility, but they are not necessary: some possibilities are not directly grounded in dispositions. I think that the major problem with this kind of solution is that one gives up the initial motivation of the modal dispositionalist: grounding every possibility in dispositions. One loses the explanatory strength of the theory.

In this paper, I want to suggest another provisional answer to the problem. I think that the classic modal dispositionalist has the resources to refute *necessitism*. An important step is to notice that the claim of the **Necessity of origin** isn't self-evident. I contend that the modal dispositionalist can argue for this **(i)** by noting that the apparent attractiveness of this claim relies on certain implicit assumptions concerning the former debate after Kripke (1980) and his affirmation of the essentiality of origin. The reference is explicit in Vance (2014) and Vetter (2015), but remains still under-exploited. I will argue **(ii)** that the modal dispositionalist could then find the proper resources to deny the necessity of origin, and suggest a precise argument for it.

2. The necessity of origin and the use of a branching model

In order to understand the claim that the origin is necessary, I argue that it is important to see how it relies on a classical debate after Kripke (1980). According to him, the origin of an object is a non-trivial essential property. This applies both to inanimate objects and living beings.

There are two different interpretations of this claim. The first interpretation of Kripke's claim asserts that the essentiality of origin presents a sufficiency principle. The origin seems then to be a necessary and sufficient condition to be a specific object. This interpretation is the most important one found in the literature and also the metaphysically heaviest. It has been supported most notably by McGinn (1976), Salmon (1981; 2005), Forbes (1985) and Noonan (1983). It is yet challenged by two major problems: the *recycling problem* and

the *tolerance problem*. What this discussion ultimately leads to is that neither the sufficiency principle nor its conclusion, the necessity of origin, is self-evident (see Robertson and Atkins (2023) for an overview).

By considering those sceptical issues, J. L. Mackie (1974) and P. Mackie (1998; 2006: chapter 6) suggest another argument for the necessity of origin which relies no longer on sufficiency principles, but on a *branching model*. This model can only allow *forward branchings* and divergences. The origin is then a necessary feature of it. One can always draw a new segment diverging from a point *p*, and this even in the past. The origin of the segment, however, is fixed by definition since there are no possible converging segment. In other words, possibilities must always start somewhere (i.e. in the origin).

Two remarks can be made here. Firstly, I want to note a defining limitation. The model has been designed explicitly to represent *de re* possibilities, and not *de dicto* ones. It seems then that it cannot directly claim to represent every metaphysical modality. Complex possibilities or logical necessities for instance do not seem to fit well into the model. P. Mackie explicitly acknowledges this restriction (2006: 107-108). Secondly, there is another limitation, which seems to restrict the scope of the model. It appears too coarse-grained and prevent more fine-grained distinctions from being made. The model shows that the origin is a necessary feature so that the various diverging possibilities can then be constructed. But it does not tell us what the origin might consist of. Worse still, since it makes the origin of the segment necessary, then all the circumstances surrounding it are at the same time necessary by the same token. She explains then that a distinction must be made between two very different questions (2006:98):

1 – Why origin (rather than development)?

2 – Why *these* features of origin (rather than the others)?

According to her, the first question is more fundamental, and this is the one the branching model addresses. The second, on the other hand, is the question concerning the sufficiency principle. What I am worried about here is not the branching model itself than its extension to other questions that are not within its scope and its claim to account for all modalities whether *de re* or *de dicto*. It is also worth noting here that if one applies the second question to the

branching model, it turns out to be equally subject to the *recycling problem* and the *tolerance problem*.

That being said, one could notice then that the modal dispositionalists almost only refer to *branching models* when arguing for the necessity of origin. They explicitly revive this post-Kripkean debate, but within a different framework. The debate is no longer localized to the essence of objects, but is considerably extended to the universe as such. One may note from the outset that the step, not taken by P. Mackie, of extending the model to all metaphysical modalities is taken.

This necessity of origin has been highlighted by Cameron (2008) and Vetter (2015: chapter 6) and happily approved by Vance (2014) and Kimpton-Nye (2021). But the main arguments for this claim are to be found especially in Vance (2014) and Vetter (2015). Their views are similar and have a number of points in common. First, they appeal to the relation of dispositions to time and then they explicitly both apply Mackie's branching model. One may already note that the reference to the sufficiency principle is almost absent, which might be seen as a significant shortcoming. I will only focus on Vetter's argument, since Vance explicitly and directly applies Mackie's *branching model*. Vetter's (2015) seeks to build a theory which would allow, while reducing all modalities to the dispositions of actual objects, to find back the *de dicto* modalities of which one can have an intuition. As one can observe, the modal dispositionalist want to account with the most acute granularity what precisely grounds the *de dicto* modalities. The branching model, on the other hand, has a more limited scope since it can only coarsely and indiscriminately represent the modalities. However, Vetter uses it in order to account for the temporal asymmetry of dispositions. Dispositions are *forward-looking*. They are dispositions for a possible future manifestation. On the other hand, the possession of past dispositions is always *trivial*. There is then a modal asymmetry in the dispositional framework due to what she calls *the triviality thesis*:

Triviality Thesis – Nothing has a disposition at some point in time t for the state of the world at any $t' \leq t$ (i.e. at any time t' before or identical to t) to be different from what it is like at t' .

The reference to the P. Mackie's "why origin" question is explicit. Vetter notes the undeniable proximity between the branching representation of possibilities and the potentiality-based one (see 2015:204, note 4; 291). There seems to be, according to Vetter's analysis (although cautious and tentative), a fundamental difference between past concerning dispositions and future concerning ones. Dispositions possessed in the past must necessarily manifest themselves, whereas future-oriented ones may or may not. Their manifestation remains contingent. This temporal asymmetry can be then also extended to account for the claim that the origin of the universe is necessary (2015:205).

Vetter (2015: section 5.8) and Kimpton-Nye (2021:14) suggest that the only likely solution to refuse the *triviality thesis* would be either backward causation or time travel. It seems to me that this double suggestion is mistaken and can only lead to dead ends.

First, concerning backward causation, even if this had important consequences for the notion of causality, it clearly misses the point of the *triviality thesis* and the necessity of origin. An important distinction must be made here. The notion of backward causation is only the idea that the temporal order of cause and effect is a mere contingent feature, and that there are cases where the cause is causally prior to its effect but temporally posterior. But the causal link is the same. The problem facing the modal dispositionalist is rather to know if the past could be different or changed. It is a completely different question. A distinction must be made between changing the past, such that it could be different from what it is, and influencing the past, such that it could only be such that it is. Backward causation entails only that the future could have an influence on what happens in the past, but this absolutely cannot change their modal status. The past, being causally influenced by the future, is just as fixed.

Time travel seems to be a better option for *changing the past*. However, it seems to me just as unable to account for the contingency of origin. First, contrary to backward causation, if a system S is travelling in time, it would preserve its temporal order during the trip, so the *triviality thesis* is still effective on this limited scale. But then suppose that I want to change the origin of the universe and I go back to the first state s_1 . The paradox is that if I

want to completely change the past, there must always be that state s_1 beyond which I cannot go. This first state, being the ultimate hypothetical point at which one can go back in time, is necessary.

It seems to me that those two solutions are thus far from less promising than they appear. In the next and final section, I will present an argument which, if successful, will at least show that the *triviality thesis* and the necessity of origin are not as robust as it seems.

3. An argument for the contingency of origin

This argument has two parts.

First, I want to question the use of Mackie's branching model in the modal dispositionalist framework. It doesn't seem fit as an explanatory and heuristic model for the dispositional grounding of *de dicto* modalities. Let's take our initial interrogation: is there some xx such that they have, had, or will have an iterated or noniterated disposition for the origin of the universe to be different? The problem is not to distinguish the origin from the future development, but rather to know whether there are dispositions such that they can ground the *de dicto* statement: "It is possible that the very first state of the universe could have been different". The question is to know if there are dispositions such that their origin could have been different. This is typically a "why these features?" question: what are the features of origin that are necessary and sufficient to give rise to those dispositions? So, one is looking precisely for what is *non-trivial* about the origin. If I am right, then it is appropriate to contest P. Mackie's hierarchy of questions, particularly for *de dicto* modalities, and reject the branching model as an inadequate explanatory model for our present problem.

The inversion of the two questions allows me, secondly, to note that this gives one the opportunity to return legitimately to the debates on the sufficiency of origin, since it is a matter of knowing what is necessary and sufficient in the origin to produce a disposition. Hence one can easily argue that modal dispositionalism isn't immune to the two objections facing the sufficiency principle, *the recycling problem* and *the tolerance problem*. Those objections are also applicable to the necessary origin within a dispositionalist framework.

The difference is that they have to be reformulated without reference to possible worlds. I will focus on the *tolerance problem* which questions not only the *sufficiency* of origin, but also its *necessity*.

I suggest what I would call the *glassmaker's example* inspired by Chandler (1976). Suppose a glass. For the modal dispositionalist it is defined by its dispositional properties, such as its specific disposition to break precisely if struck with a force from 8 N. Suppose then that this dispositional essence is tolerant: it may vary precisely between 7.5 N and 8.5 N, but no more than that. So, the tolerance margin is very precise and not vague here. For a variation to happen, the glassmaker should slightly change the fabrication of the glass, by using a different shape or a change in the composition of the glass. The glass's disposition to break is then such that it could ground the possibility that its shape or composition could have been different. This being admitted, it is possible to use S4 and the transitivity of iterated dispositions to prove the contingency of origin. There is a discussion about whether modal dispositionalism could ground the S5 axiom (see Vetter (2015: section 6.4), Kimpton-Nye (2021) and Werner (2022b)). My argument is neutral on this issue. S4 is quite independent of S5 and seems to fit much better with modal dispositionalism and the notion of iterated dispositions. S4 says that if x is possibly possible, then x is possible. Within the modal dispositionalist framework, this gives: if some xx have an iterated disposition for some yy to have an iterated disposition that f , then the xx have an iterated disposition for f . This appears plausible. So, returning to my example and given the transitivity of disposition, one could argue that if the glass has a disposition such that it could have had a different origin (for instance the glassmaker could have manufactured the glass differently) and in turn an iterated iterated disposition such that the glassmaker could have learned differently how to manufacture glasses, and so on. One can then suppose a hypothetical chain of iterated dispositions such that the origin of the universe could have been completely different from what it is. The glass is then such that there is the possibility p that $\neg H$, that the origin of the universe could have been different. An ordinary disposition then such as a glass's disposition to break could be argued to ground the possibility that the origin of the universe could have been different.

4. Conclusion

Given the limitations of the *branching model* and its use for the argument of the necessity of origin, I attempted to present a possible ground for the contingency of origin relying on the *tolerance principle*. One could then find the resources to argue that the claim of the necessity of origin appears superfluous. My conclusion here is that this solution seems to reject *necessary perfect masks*. It is a form of *No Mask* solution (see Vetter and Busse (2022)). The glass that never breaks only *contingently* never breaks. This disposition is compatible with a specific tolerant set $H = \{h_1, h_2, h_3 \dots h_n\}$ of possible origins that all could have led to the same disposition and possible manifestations.

References

- Borghini, Andrea, and Neil E Williams (2008) "A Dispositional Theory of Possibility", *Dialectica* 62 (1), 21–41.
- Cameron, Ross (2008) "Truthmakers and Modality", *Synthese* 164 (2), 261–80.
- Chandler, Hugh S. (1976) "Plantinga and the Contingently Possible", *Analysis* 36 (2), 106–9.
- Forbes, Graeme (1985) *The Metaphysics of Modality*, Oxford: Oxford University Press.
- Jacobs, Jonathan D. (2010) "A Powers Theory of Modality - or, How I Learned to Stop Worrying and Reject Possible Worlds", *Philosophical Studies* 151, 227–48.
- Kimpton-Nye, Samuel. (2021) "Can Hardcore Actualism Validate S5?", *Philosophy and Phenomenological Research* 102 (2), 342–58.
- Kripke, Saul (1980) *Naming and Necessity*, Cambridge, Mass.: Harvard University Press.
- Mackie, J. L. (1974) "De What Re Is de Re Modality?", *The Journal of Philosophy* 71 (16), 551.
- Mackie, Penelope (1998) "Identity, Time, and Necessity", *Proceedings of the Aristotelian Society* 98 (1), 59–78.
- Mackie, Penelope (2006) *How Things Might Have Been: Individuals, Kinds, and Essential Properties*, Oxford: Oxford University Press.
- McGinn, Colin (1976) "On the Necessity of Origin", *Journal of Philosophy* 73, 127–35.
- Noonan, Harold (1983) "The Necessity of Origin", *Mind* 92: 1–20.
- Robertson, Teresa, and Philip Atkins (2023) "Essential vs. Accidental Properties", *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta and Uri Nodelman, Spring 2023 edition. <<https://plato.stanford.edu/archives/spr2023/entries/essential-accidental/>>.
- Salmon, Nathan (1981) *Reference and Essence*, Princeton, NJ: Princeton University Press.

Salmon, Nathan (2005) *Reference and Essence*, 2nd edition with added appendices. Amherst, NY: Prometheus Books.

Vance, Chad (2014) "Dispositional Modal Truthmakers and the Necessary Origin", *Philosophia* 42 (4), 1111–27.

Vetter, Barbara (2015) *Potentiality: From Dispositions to Modality*, first edition. Oxford Philosophical Monographs. New York, NY: Oxford University Press. Vetter, Barbara, and Ralf Busse (2022) "Modal Dispositionalism and Necessary Perfect Masks", *Analysis* 82 (1), 84–94.

Werner, Jonas (2022a) "Extended Dispositionalism and Determinism", *Ergo an Open Access Journal of Philosophy* 9 (0).

Werner, Jonas (2022b) "Potentialism and S5", *Canadian Journal of Philosophy* 52 (6), 622–35.

Acknowledgement

I wish to thank Jean-Pascal Anfray, Adrien Avramoglou, Lucas Escobar, Mathilde Escudero, Rayan Geha, Jiaqi Li, Louis Pijaudier-Cabot, Sophie Roux and Stéphanie Ruphy, the members of the Science and Metaphysics research seminar (ENS, Paris), the members of the Mathesis group (République des Savoires, UAR 3608) and the members of "Langage, Epistémologie et Métaphysique" (ENS-SU, Paris) for the very helpful discussions and feedbacks.

Being Queer

Matthew Andler (Madrid, Spain)

Abstract

Extended Abstract. Here, I provide a general overview of an in-progress book, tentatively titled *Sexuality in Social Context: A Philosophical Analysis*. The book aims to develop a systematic social constructionist theory of sexual orientation and sexual identity.

In “What’s Special About Sexual Orientation?,” I ask what (if anything) might meaningfully distinguish sexual orientations such as *asexual*, *bisexual*, *homosexual*, and *heterosexual* from other aspects of human sexuality such as sexual dispositions with respect to height, weight, or power dynamics. I argue that the dispositions that ground sexual orientations are not intrinsically differentiated from dispositions that do not ground sexual orientations. Instead, I argue that sexual orientations are only special in virtue of having significance within (what I call) *heteropatriarchal kinship structures*, which are key to arguments later in the book (as well as to new work on friendship and love). On my view, heteropatriarchal kinship structures unjustly organize sex, love, reproduction, and care via ideologies that privilege relationships in which sex, love, reproduction, and care are organized “wholesale” within dyads between cisgender women and cisgender men.

Next, in “Sexuality is Essentially Sociopolitical,” I critique recent “gender-critical feminist” arguments that aim to exclude trans women from queer spaces. The aforementioned arguments appeal to trans-exclusive interpretations of sexuality categories such as *lesbian*, which in turn rely on (what might be called) “common sense epistemology.” While I agree that philosophy must start somewhere, such that ordinary judgements in ordinary contexts might be taken into account as part of metaphysical theorizing, I argue against the idea that trans-exclusive interpretations of sexuality categories amount to common sense in any epistemically authoritative way. Given that common sense is variegated, with incompatible judgments, any work in social ontology that appeals to common sense must answer a methodological question: how ought a theory to decide from which ordinary judgements to take guidance? At minimum, I expect that gender-critical feminists would agree that the common sense that would be epistemically authoritative is not the common sense to be found in contexts of gender and sexuality oppression. Instead, if common sense is epistemically authoritative, it is the common sense to be found in liberatory social movements. But the ordinary judgments in the aforementioned contexts are (at minimum) indeterminate and (more likely) trans-inclusive, such that any methodologically defensible appeal to common sense would not provide reason in favor of gender-critical feminist arguments aim to exclude trans women from queer spaces.

In “Being Queer,” I defend the cultural analysis of sexual identity as originally developed in “Queer and Straight” (2022, in: Clare Chambers, et al., eds., *The Routledge Handbook of the Philosophy of Sex*, New York: Routledge, 117-30). Foundational to this work is a distinction between sexual *orientation* and sexual *identity*, in which “sexual identity is the social meaning of sexual orientation” (M. Andler, 2021, “The Sexual Orientation/Identity Distinction,” *Hypatia* 36, 259-75). Here is the cultural analysis of queer sexual identity: “An individual has a queer sexual identity in virtue of (i) being excluded from straight culture and (ii) being such that according to the constitutive norms of queer culture the individual ought to be included in

queer culture” (ibid., 124). Likewise here is the cultural analysis of straight sexual identity: “An individual has a straight sexual identity in virtue of failing to satisfy conditions (i) and/or (ii)” (ibid.). The constitutive norms of queer culture include those related to self-constitution and solidarity; for example: “[i]f an individual is excluded from straight culture on the basis of their sexual orientation, then – according to the norm of solidarity – the individual ought to have special access to queer cultural practices that curate normatively important resources” (ibid., 126). The overarching idea here is that queer cultures are governed by social norms that determine the membership conditions of sexual identity categories such as *queer*.

Here, in response to recent criticism from Ben Caplan, I appeal to hyperintensional metaphysics in order to argue against orientation-based views of sexual identity (according to which facts about individuals being queer are grounded in facts about individuals being non-heterosexual). And in response to recent criticism from Raja Halwani, I argue in favor of the following theses. First, queer and straight sexuality cultures admit of metaphysical analysis, with cultural practices such as drag and vogue literally composing queer cultures. Second, cultural practices are queer and/or straight in virtue of resisting and/or entrenching heteropatriarchal kinship structures (as theorized in “What’s Special about Sexual Orientation?”). With these addendums, I continue to endorse the cultural analysis of sexual identity.

How Hegel brought Brandom back to Wittgenstein

Pavel Arazim (Prague, Czechia)

Abstract

Brandom is one of the most influential philosophical heirs of Wittgenstein. While accentuating his indebtedness, he also does not conceal that he reads Wittgenstein somewhat selectively, just as Hegel, another of his major inspirations. The basic lessons Brandom took from Wittgenstein are that language is based on rules and that some rules are implicit. I argue, though, that these lessons become seriously distorted when isolated from the rest of Wittgenstein's thought. Basically, this boils down to appreciating that Wittgenstein did not want to advance philosophical theories of language and had strong reasons for that. The way in which Brandom pictures the functioning of rules and how we can allegedly make them explicit, underestimates their elusiveness. We are too intertwined with our rules in order to be able to just examine and change them at our will. Despite this criticism, though, I point to the way in which Brandom, in his reading of Hegel, came to appreciate the historicity of rules. This appreciation is very close to the Wittgensteinian insight that our rules are founded in our form of life. Though he himself does not acknowledge it, there is an early and later Brandom. And the later Brandom is in a much better position to do justice to Wittgenstein, especially the later Wittgenstein. Given the influence of Brandom, it is important to note this break in his thought. It prevents overlooking what is valuable in Wittgenstein.

Besides his teacher Wilfrid Sellars together with the classics of German idealism Kant and Hegel, Robert Brandom typically emphasizes his indebtedness to Wittgenstein. Brandom's account of language and normativity develops many of great Wittgensteinian insights and Brandom thus can be seen as his philosophical heir in many respects. I would, nevertheless, want to point to some essential aspects of Wittgenstein which Brandom leaves aside. This omission has grave consequences for his philosophy. Nevertheless, I also argue that later in his development Brandom manages to save a lot of what is valuable in Wittgenstein, though he does so rather by engaging with Hegel than with Wittgenstein.

1. Inferentialism and logical expressivism or the early Brandom

The basic tenet of Brandom's approach, known as inferentialism, is that meaning is constituted by inference rules. Primarily, a sentence is meaningful due what it is entailed by and by what it entails, possibly with further premisses. Continuing the work of Sellars (1948), Brandom (1994) points out that even what comes close to the protocolar sentences which the logical positivists were so keen to find, has to be able to play a role in inferences. To

cite one of his favourite examples, when I claim that it is raining, then my utterance is meaningful partly because one can, in normal circumstances, infer from it that the streets will be wet (Brandom 2000, p.52).

Brandom tracks this normative account of meaning back to Kant, yet he sees Wittgenstein as a vital ally in this respect, too. Nevertheless, it is another ingredient of Brandom's philosophy which he considers as specifically Wittgensteinian. To start somewhat vaguely, the rules which are supposed to constitute meaning, are not easily available. Brandom speaks of the rules which are implicit. Here he has in mind those passages in Wittgenstein where he shows how we struggle to clearly formulate what the specific shape of the rules is. Very often, this concerns the rules on a metalevel which are supposed to explain some rules which we encounter more directly, for example the rules for the interpretation of the sign-post, or the rules which specify how a given order, for example continue this number sequence, is to be interpreted (Wittgenstein 1984, §85, §185).

According to Brandom (2000, chapter 1), the implicit can be rendered explicit. He even reserves for our purported capacity to do so the noble title of Socratic rationality. To illustrate, we can be implicitly guided by the rule that when it rains, then the streets are wet. Nevertheless, we may render this rule explicit and claim that if it rains, then the streets are wet. By doing so, we are able to discuss and potentially criticize and modify the given rule. Typically, this would happen in cases when a rule or a set of rules does not satisfy us in some way. Should we find the streets as good as dry after a significant portion of rains, then there seems to be something wrong with the rule and it may be the right time to change it.

Instrument for the performance of this Socratic, or also expressive rationality, is logic. We could see the conditional *if, then* at work making explicit the link between raining and wet streets. But here I think we should pause to see to what degree this picture is indeed such a straightforward development of Wittgenstein as Brandom suggests. Not only will I oppose this understanding of Wittgenstein, I will show that it is significantly less convincing than Wittgenstein himself and fails to take some important lessons from his thought.

2. The first thing which Brandom missed in Wittgenstein

There are at least two ways in which Brandom breaks with Wittgenstein. He does recognize both these ways but insufficiently. This is even more the case in the second than in the first way.

The first way of breaking with Wittgenstein consists in the fact that Brandom obviously creates a monumental theory of rules, namely inferentialism. Maybe it is closer to some fundamental aspects of the late-Wittgensteinian spirit than what Brandom calls representationalism (Brandom 2000, Chapter 1). Indeed, representationalism, the thesis that a meaning of a given word consists in what it denotes or represents, is very close to the Augustinian picture of language that Wittgenstein attacks in *Philosophical investigations*. It is clear that Brandom very much likes what Wittgenstein has to say about the philosophically-laden memories of the early childhood of the saint. Indeed, the understanding of language which is present in the text of the *Confessions*, according to which meaning is conveyed by the means of pointing to the purported referent, seems as a germ of a theory of representationalism which Brandom sees as the semantic mainstream which he wants to oppose his inferentialism to. But Brandom is still too hasty if he considers Wittgenstein as an ally here. After all, Wittgenstein opposed advancing any philosophical theory at all.

Indeed Wittgenstein (Wittgenstein, §128) claims that philosophy should not present any theses, which presumably should concern also the thesis that meaning is constituted by inference rules. Furthermore, the passage which many philosophers love to quote so much, namely §43, is rather cautious about the declaration that meaning is use. And in that very paragraph, after what became a popular slogan, Wittgenstein mentions that sometimes, meaning is clarified by pointing to a referent. Rather than an announcement of the grand new philosophical school which with the slogan that meaning is use, it is a very cautious suggestion of how we can liberate ourselves from a certain deluding image of language.

In such a context it is very surprising that one would engage in a programme of explaining reference away in order to give inference rules the opportunity to explain linguistics meaning in all its aspects, including reference, the abandoned explainer. Brandom might be free to diverge from Wittgenstein

and try proposing a theory. Yet I believe that he differs from Wittgenstein also very much in his understanding of what a rule is. Wittgenstein dedicates a lot of space to the notion of a rule, which naturally leads the reader to suppose that he deems it very important or even central. But the way Wittgenstein speaks of it might also suggest that he rather wants to dissuade us from creating theories which are based on the notion of rule. For example the passages about the sequence of even numbers can be read as an ironic hint at how little of our lives and practices can be explained by the notion of a rule. All the popular rule following considerations from the *Investigations* might be understood as warnings not to replace representationalism by something resembling Brandomian inferentialism.

3. The second thing which Brandom missed in Wittgenstein

The different understanding of rules brings us to the second way in which Brandom breaks with Wittgenstein. This divergence consists in Brandom's notion of an implicit rule. The idea of an implicit rule which can be rendered explicit may not be completely alien to Wittgenstein but still the way Brandom puts it is just too simplistic and crude. At least by Wittgensteinian lights. It is of prime importance for Wittgenstein that we are always too much entangled in rules in order to have a clear overview. Obviously, Wittgenstein seems to be trying to shed some light on rules in his writings, so perhaps he believes that there is a sense in which they can be made explicit. Nevertheless, there is also a sense in which it is absurd that we can just bring a hidden rule to the surface and decide whether we want it or not. In an important sense, we are the rules, we do not float above them. And though there are ways in which various language games can be relatively enclosed and independent of each other, there are also ways in which they are intertwined. Wittgenstein indeed insisted not only on the plurality of language games, but also on their interconnectedness, as in the end of §7, he claims that even the whole of language with related activities should be considered as a language game. A given rule, then, is in some sense what it is only in the context of all other rules, indeed of the whole form of life that we are. If one rule is supposed to be implicit, then all rules are. Maybe it makes sense to say that some rules are less

implicit than others, but implicit nonetheless. Therefore the very distinction between explicit and implicit rules on which Brandom relies, is highly problematic.

Brandom is partly aware of this difference between his account and Wittgenstein's when he discusses whether the language has a downtown. For him, the downtown, the core of the imaginary city of language, is the game of giving and asking for reasons. Brandom (2000, pp. 14-15) denotes his approach as rationalist. For him, arguing 'seriously', so to say, is necessary in order to engage in any other linguistic activity. We can compose poetry, tell jokes, play with words in many ways but all these suburban activities are genuinely linguistic only when we engage in the central game of giving and asking for reasons. It seems that the scenario in which some people would use language, for instance, merely to tell jokes, does not make sense for Brandom. He would contest that what they would engage in truly linguistic activities. It would not make sense to interpret the sounds issuing from their mouths as words which conspire to make funny sentences. Wittgenstein would likely agree but he would also be ready to argue the other way round. It hardly makes sense to countenance serious argumentation without countenancing jokes and all the other things we do with words, to use the phrase of Austin (1962).

What is the picture of the city of language that Wittgenstein conveys to us? In a way, it seems quite opposite to what Brandom confronts us with. Wittgenstein (Wittgenstein 1984, §18) speaks of a more typically European city such as Vienna, Prague, Paris and the like. In the centre we have curvy and somewhat chaotic streets, in the suburbs we have rather a more linear and clearly organized districts. Brandom seems to see rather the center as rectilinear and the suburbs as whimsical, curvy and also perhaps more entertaining and charismatic.

In fact, when Brandom claims that some rules are implicit in practice and can be made explicit by means of inference rules, he renders himself vulnerable to the criticism Wittgenstein addressed to Augustine. Brandom presents an image of language practice which is too discursive. Wittgenstein did not want to see the thoughts of a child Augustine as a kind of language which only awaits its

expressions. Equally, I believe, we can criticize the notion of your practices purporting to say something, namely to state rules, which only await an explicit expression.

The notion of rule Brandom had was therewith too intellectualist for Wittgenstein and I believe that we can see in what way it was indeed problematic. We can observe our practices or practices of others and it can be helpful to formulate rules which we observe in them. But that does not mean that there are always specific rules which are correctly associated with a given practice. Wittgenstein could use his early notions of *Zeigen* as opposed to *Sagen* in order to differentiate himself from Brandom. We can only point to the language games we play, we cannot make them explicit in the way Brandom pictures. We are just too much inside our form of life in order to be able to express it. This does not imply that there could be a divine point of view from which the true shape of our rules can be glanced at, rather it is meant to show that it makes no sense to strive for such a correct expression. There is no correct expression.

4. The Hegelian turn

But Brandom has moments where he seems to realize the problematic aspects of his conception and indeed become more Wittgensteinian. This is the case for example in first chapter of *Articulating reasons* where he speaks of how the inference rules are modified and claims that Dummett's notion of harmony between inference rules gets its content only in the course of work with the concepts we have (Brandom 2000, p.75). We do not have a notion of harmony in advance and then only apply it, we have it only as we work on it. This reflection, I believe, could be applied to very notion of rule and of an expression of rule, which are so central for inferentialism.

But it is only in his relatively recent *Spirit of trust* that Brandom (2019) truly manages to correct the shortcomings of his early approach. In this book, Brandom makes explicit his hitherto rather implicit philosophical passion for Hegel. In the course of interpreting the *Phenomenology of Spirit*, Brandom shows that our concepts and our knowledge are radically historical and

depend on the specific context in which they have arisen. This Hegelization ultimately brings him closer to Wittgenstein and makes him partly abandon the too naive logical expressivism which he defended earlier.

Brandom differentiates, following Hegel and the tradition of German idealism, between understanding, in German *Verstand*, and reason or *Vernunft*. While understanding considers its concepts as given and static, reason understands that they are dynamic in their nature. Creating them and applying them cannot be really separated. Brandom (2019, p. 17) writes that it is of essence of *Vernunft* that it regards social institutions both as created and as discovered, that it reckons with both these perspectives. And I believe on the passage from his early to his later philosophy, Brandom himself passed from *Verstand* to *Vernunft* which is also a more Wittgensteinian perspective.

Both Brandom and Hegel still probably are much more optimistic about the possibility of expressing the rules which make up the fabric of our society than Wittgenstein. But the later, Hegelian Brandom is certainly closer to Wittgenstein than the early Brandom. Rules are not anymore simply here for us to inspect them and make them explicit, partly because they are still in the making. Applying the old rules means also re-creating them, giving them a new shape.

Just as there are significant foreshadows of the later Wittgenstein in the *Tractatus*, i.e., in the early Wittgenstein, so there are signs of the specifically later Brandom in the early Brandom. Nevertheless, it took an irreducible leap to make the change.

The historicity of rules prevents us from taking the rules lightly out of their historical situation and asking how they could work in a different context. Because the creation and discovery of the rules are intertwined, seeing the rules just as highly specific entities which are simply here for us to inspect, is a picture which Brandom overcomes. And it does not matter so much whether these rules are thought of as hidden under some surface as implicit rules or whether they are in the open and known as the explicit rules.

Brandom, following Hegel, interprets the rules as necessarily historical in the sense that only if we understand them as a result of an error, as a corrected error, can we overcome scepticism, the topic Wittgenstein had targeted since

the *Tractatus* and which is central especially in *On certainty* (Wittgenstein 1969). While Brandom focuses on the scepticism claiming that we cannot know the reality as it is, his argument works even against the scepticism which casts its shadow over the rules in *Philosophical investigations*, which is the scepticism that fascinated Kripke (1982) so much.

While Brandom and Hegel are more keen on constructing broad visions, whereas Wittgenstein remains much more terse and indeed tuciturn, we can see that the later Brandom not only is closer to Wittgenstein but might also cast a useful light on how our rules are in a way not ours, as they are always somewhat inaccessible. This is what he, I believe, was aiming at, though in a flawed manner, when he spoke of the implicit rules. While probably not entirely corresponding to what Wittgenstein was aiming at, just as he in many ways may misread many fascinating passages in Hegel, Brandom does a remarkable job at pointing to the insights which are not so easy unearth in the writings of both these difficult philosophers. There might be a sense in which Hegel still is closer to Wittgenstein than Brandom is, namely by considering the rules much more as embodied in specific historical situations and much less as abstract entities which can or cannot be realized in those specific contexts. That Hegel and Wittgenstein are much closer than one might suspect is suggested, among others, by Kolman (2019). But Brandom manages to revive a lot of what could otherwise remain unnoticed.

The rules can sometimes be much more in our power than we might dream of. At other times, though, they can be much less under our control than we can imagine and the idea of making them explicit, let alone of modifying them in any way, is naive. These are great insights of Wittgenstein. And Brandom, with the help of Hegel, has opened them for himself and the contemporary philosophy of language. It took him a little bit longer to get to the second insight but that is a minor issue.

References

Austin, John L., (1962), *How to do things with words*, Oxford: Oxford Univesity Press.

Brandom, Robert (1994) *Making it explicit*, Cambridge MA: Harvard University Press.

Brandom, Robert (2000) *Articulating reasons*, Cambridge MA: Harvard University Press.

Hegel, G.W.F. (1807) *Phänomenologie des Geistes*, Bamberg/Würzburg: Verlag Joseph Anton Goebhardt .

Kolman, Vojtěch (2019) “Master, Slave and Wittgenstein: The Dialectic of Rule-Following”, in: Jakub Mácha and Alexander Berg (eds.) *Hegel and Wittgenstein*, Berlin: de Gruyter, 227-242.

Kripke, Saul (1982) *Wittgenstein on Rules and Private Language*, Cambridge MA: Harvard University Press.

Sellars, Wilfrid (1948) “Concepts as involving laws and incoceivable without them”, *Philosophy of Science*, 15, 287-313.

Wittgenstein, Ludwig (1969) *On Certainty*, Oxford: Blackwell, translated by Denis Paul and G.E.M. Anscombe.

Wittgenstein, Ludwig (1984) *Werkausgabe Band 1. Tractatus logico-philosophicus / Tagebücher 1914–1916 / Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp.

Acknowledgement

This work was supported by from the grant project of The Czech Science Foundation (GAČR), number 23-07119S with the title *Meaning as an object - principles of semantic theories*. The project is led by Jaroslav Peregrin, my colleague from the Institute of Philosophy of the Czech Academy of Sciences.

"Will the Sun Rise Tomorrow?" Wittgenstein on Scepticism and Hinge Propositions

Lorenzo Barba (Chieri, Italy)

Abstract

In this paper I deal with the problem of 'hinge propositions' in Wittgenstein's later writings, and in particular in the manuscript published after his death under the title *On Certainty* (OC): here he states that there are propositions which we cannot call into question and which appear to be certain in the highest degree (OC 341, 342). Hence, the interpretation of this writing is particularly important in the contemporary debate on the foundations of knowledge.

I will try to explain the features of 'hinge propositions' through a brief analysis of the origins of this concept. To do this, I'll first consider the problem of elementary propositions in his *Tractatus Logicus Philosophicus* (TLP), with the aim of showing its significance in relation to Wittgenstein's conception of causal laws. Then I'll try to explore the development of these themes in his later writings. In particular, I will focus on the claim he made in the *Tractatus* that we cannot know that the sun will rise tomorrow (TLP 6.36311). Using this example as a starting point, I will try to understand the development of Wittgenstein's thought on this subject: this analysis will show that hinge propositions play a semantic role, that is, a role concerning the meaning of our sentences.

The independence of elementary proportions and the existence of laws of nature

In his early writings, Wittgenstein assumes that semantics must be independent of ontology, which means that the meaning of a proposition must not depend on the truth of another proposition, as he remarks in a note to Moore:

"The question whether a proposition has sense (Sinn) can never depend on the truth of another proposition about a constituent of the first." (Wittgenstein 1961: 18)

Indeed, if the meaning of a proposition (e.g. "p") depended on the truth of another proposition (e.g. "q"), the meaning of the second proposition would in turn depend on the truth of a third proposition, and this would generate a regressum ad infinitum (see also Frascolla: 2000, p.91). Prima facie, this principle is far from obvious: if we consider, for example, the sentence 'the present king of France is bald', the meaning of the sentence depends on what we mean by 'king of France', and so we probably have to assume that the sentence 'there is a king in France' is true, in line with Russell's well-known

solution (Russell 1918: 125; on this point see also Diamond 1996: 73-93). The Tractarian analysis of language is designed to avoid this kind of problem (see Barba 2023: 42-43); indeed, thanks to the use of truth-tables, it is possible to make the logical relations between propositions clear, so that the meaning of molecular propositions clearly depends on the meaning of elementary propositions (TLP 4.4). Thus, if we consider an ideal language, each elementary proposition is independent of the others and "every proposition is a truth-function of elementary propositions" (McGinn 2022: 115, see also Gargani 1993: 31).

This idea has some important consequences for epistemology, since, for example, the truth of the proposition 'the sun has risen today' cannot imply the truth of the proposition 'the sun will rise tomorrow' (TLP 6.36311). As he explains:

"A necessity for one thing to happen because another has happened does not exist. There is only logical necessity." (TLP 6.37)

In fact, the hypothesis that the sun will rise tomorrow depends on a more general hypothesis represented by the truth of Kepler's laws (on the basis of which we can predict not only that the sun will rise tomorrow, but also what time it will rise). Hence, our prediction cannot be more than hypothetical:

"It is an hypothesis that the sun will rise tomorrow: and this means that we do not know whether it will rise." (TLP 6.36311)

The consequence of the independence of elementary proportions is, therefore, that it is impossible to formulate an a priori principle affirming the existence of natural laws:

"We cannot infer the events of the future from those of the present. Belief in the causal nexus is superstition." (TLP 5.1361)

"At the basis of the whole modern view of the world lies the illusion that the so-called laws of nature are the explanations of natural phenomena." (TLP 6.371)

However, this kind of impossibility does not lead to a form of scepticism; indeed, even if we do not know whether the sun will rise tomorrow, we do not

doubt that it will. The lack of knowledge in this case does not imply a lack of certainty or a state of doubt.

"If there were a law of causality, it might run: 《There are natural laws》 . But that can clearly not be said: it shows itself." (TLP 6.36, see also TLP 6.32)

The possibility of natural laws cannot be proven, but it lies unexpressed at the basis of our knowledge. It is for this reason that scepticism is considered not to be false, but "nonsensical":

"Scepticism is not irrefutable, but obviously nonsensical, when it tries to raise doubts where no questions can be asked." (TLP 6.51)

The mistake of metaphysicians, from this point of view, is to try to answer a question we cannot either ask.

The revision of the Tractatus and the notion of 'grammar'

When Wittgenstein started revising the Tractatus, he began to doubt the possibility of creating a perfect sign language. In a writing named *Some Remarks on Logical Form* (RLF) we find an important step in the revision process. There he considers sentences containing numbers expressing the degree of a value:

"If someone asks us 'What is the temperature outside?' and we said 'Eighty degrees', and now he were to ask us again, 'And is it ninety degrees?' we should answer, 'I told you it was eighty' We take the statement of a degree (of temperature, for instance) to be a complete description which needs no supplementation." (RLF: 167)

His argument suggests that the sentence "The temperature outside is eighty degrees" logically contradicts the sentence "The temperature outside is not ninety degrees". On the basis of the tractarian conception of language, we should conclude that these sentences aren't elementary propositions, but rather molecular ones that require further analysis:

"One might think--and I thought so not long ago -that a statement expressing the degree of a quality could be analyzed into a logical product of single statements of quantity and a completing supplementary

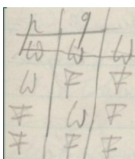
statement. As I could describe the contents of my pocket by saying " It contains a penny, a shilling, two keys, and nothing else ." (RLF: 167)

However, there are some problems with the analysis of these propositions, and in particular the use of a symbol for equivalence is problematic; Wittgenstein explains this point with reference to the concept of brightness:

"For let us call the unit of, say, brightness b and let $E(b)$ be the statement that the entity E possesses this brightness, then the proposition $E(2b)$, which says that E has two degrees of brightness, should be analyzable into the logical product $E(b) \& E(b)$, but this is equal to $E(b)$; if, on the other hand, we try to distinguish between the units and consequently write $E(2b) = E(b') \& E(b'')$, we assume two different units of brightness; and then, if an entity possesses one unit, the question could arise, which of the two- b' or b'' - it is; which is obviously absurd." (RLF: 167)

Even if he was not completely satisfied with this argument and he decided not to present the writing to the public, however he went on working on similar ideas, as it is testified by his conversations with Ramsey and Sraffa. The conclusions he reached are essential to the later concept of "grammar": he acknowledged that two elementary proportions may exclude each other, and this means that there are linguistic norms which cannot be expressed by the rules of formal logic (McGinn 2022: 123). In a manuscript, he illustrates this by deleting a line in the truth table of the logical product between the two propositions (Wittgenstein 1994 Band 1: 58):

p	q	\mathbb{W}
\mathbb{W}	\mathbb{W}	\mathbb{W}
\mathbb{W}	F	F
F	\mathbb{W}	F
F	F	F



The image shows a handwritten truth table on aged paper. It has three columns labeled 'p', 'q', and 'W'. The rows represent combinations of truth values: (W, W), (W, F), (F, W), and (F, F). The second row, (W, F), has a horizontal line drawn through it, indicating its deletion. The other rows are intact.

(Fig.1, Wittgenstein's original scheme is on the right)

The concept is that certain sentences, even if they are not contradictory from a strict logical point of view, must be excluded from language as meaningless, as he would later write in the *Philosophical Investigations* (encl. PI, see PI 500). The conclusion is that sentences like "If the temperature is 80 degrees outside, it is not 90 degrees" have a grammatical rather than an epistemological function, inasmuch they are the articulation of the

comprehensive system of rules that sustains our language. The significance of this new perspective on language is also relevant to the theory of knowledge.

Scepticism and hinge propositions

A manuscript, which has been published with the title *On Certainty*, discusses Moore's arguments against idealism. These arguments were presented in a lecture held by Moore in 1939 at the British Academy: in essence, Moore asserted that there are fundamental statements which are evident and true for all individuals and which form the foundation of our knowledge. These statements can be referred to as 'truisms' and represent common-sense truths, as Baldwin (2010) states. In summary, the author aimed to defend a common sense realist perspective and to invalidate idealist views that deny the existence of an external world. To support his argument, while replying to his critics, he referred to the presence of material objects using his own hand as an example and stating:

"This hand is a material thing; therefore there is at least one material thing." (Moore 1942: 668).

As Coliva points out, a modern exposition of the argument might be the following (Coliva 2012: 13):

1. This is my hand;
2. If there is a hand here, then the external world does exist;
3. The external world does exist

The sentence (2) is equivalent to the sentence "if the external world does not exist, this is not my hand". In this way, according to Moore, it is possible to prove the existence of the external world.

In contrast to Moore's view, Wittgenstein's remarks aim to show that propositions which seem to represent the highest level of epistemic certainty are in fact grammatical propositions (see Conant 1998: 249-250). He analysed Moore's argument and pointed out that it only makes sense to discuss knowledge if it makes sense to doubt (OC 622), and he maintains that a

differentiation can be drawn between sceptical doubts concerning the external world and the commonplace inquiries we pose about the objects surrounding us (Williams 2004: 143, McGinn 2022: 11). When sceptical doubts are raised, with McGinn's words, about "facts which are, on a particular occasion, available to anyone with mastery of the appropriate concepts" (McGinn 2022: 15), the result is the invalidation of the question and of any effort to answer it. Wittgenstein's comments are unambiguous on this point:

"I know that a sick man is lying here? Nonsense! I am sitting at his bedside, I am looking attentively into his face. - So I don't know, then, that there is a sick man lying here? Neither the question nor the assertion makes sense. (OC 10)

Similarly, we can draw a comparison between the case of the possible existence of a planet, say Saturn, where doubts might arise, and the question the existence of my hand, or that of the external world (OC 52, OC 20 and see also McGinn 2022: 12). In other words, the issue with sceptical doubt does not lie in the query itself, but in the conditions in which it is posed. As Conant points out, the sceptic faces a dilemma because he can express doubts, which nevertheless are not "the sort of super-doubt that he is after", or he can strip his words from the usual language-game to express 'super-doubts', but in this case he "remains unclear which of the many things he can mean by his words he wants to mean" (Conant 1998: 250).

To illustrate this point, we can consider the previous example: does the sentence "The sun will rise tomorrow" express real knowledge? In other words, we can ask what it means for someone to doubt that the sun will rise tomorrow. Is it possible to doubt this? Interestingly enough, it is possible to find an example in Wittgenstein's diaries, where here is a note in which he wondered whether the sun would have risen the next day. The odd thing is that when he posed the question he was not interested in its possible philosophical implications; he was in Norway, it was March (14.3.1937), and he was wondering whether he would have seen the sun the next day:

"Ich glaube, daß heute die Sonne in mein Fenster hereinscheinen wird." (MS-183, 213[2])

"I think the sun will shine through my window today"

And later he added:

"Bin wieder enttäuscht worden." (MS-183, 213[2])

"I've been disappointed again"

To be precise, he was near Skjolden (latitude 61°N, more or less), and here the phenomenon of the polar night is not complete: the sun was probably only a few degrees above the horizon, and it was hidden by the mountains to the east. The point is that in the right context (say, near the North Pole or near the South Pole) the question is not a philosophical problem at all. This means that it makes sense to say that we 'know', 'doubt', 'wonder if', 'ask', 'be sure', 'be certain or uncertain' if, we are able to make a projection of meaning in an appropriate context (Conant 1998: 239-241). Wittgenstein explains the question in this way:

"Just as the words 'I am here' have a meaning only in certain contexts, and not when I say them to someone who is sitting in front of me and sees me clearly, – and not because they are superfluous, but because their meaning is not determined by the situation, yet stands in need of such determination." (OC 348, see also Conant 1998: 240)

From this point of view, it is possible to question the time of sunrise on a given day and, for example, I can say that I know that the sun rise on Monday 12 August 2024 exactly 43'52" after 5 a.m. in Vienna; someone can say that he 'doubts' that this will be the case, or that he 'believes' it. However, if I asked whether the sun will rise tomorrow, the meaning of the question would not be clear: that is, it would not be clear whether I thought I'm near the Pole (in August the question is relevant near the South Pole). McGinn came to similar conclusions when she stated that "within our ordinary practice, we cannot doubt" apart from ordinary doubts (McGinn 2022: 10-11), and, in her opinion, it is possible to dismiss sceptical doubts because "we are in some sense justified in not doubting" (McGinn 2022: 10). Nevertheless, in these situations it is questionable if we are entitled to say we know it, or we are certainly about it: the point is that the meaning of these sentences is not clear (Conant 1998: 249). Similar remarks can be made concerning the names of colours:

"If someone were to look at an English pillar-box and say "I am sure that it's red", we should have to suppose that he was colour-blind, or believe he had no mastery of English and knew the correct name for the colour in some other language. If neither was the case we should not quite understand him. (OC 526).

The question of the existence of "material" or "physical objects" may be perceived as similar too. We can inquire about the objects inside a box or we can ask about presence of material objects in the Universe a few milliseconds after the Big Bang (say in the Plank era, when the Universe was 10–43 s old), and the answer could be negative, according recent theories (Ridpath 2018). However, if one were to ask about the existence of "physical objects" in the present world, the response would likely consist in an explanation of what is the meaning of the word "object" (see OC 36). Thus, it is possible to talk about 'knowing' concerning sentences like these, but it is a kind of knowledge that is different from empirical knowledge, as McGinn explains (McGinn 2022: 62). In this sense, on the other hand, it would be possible to doubt these sentences for someone who speaks poor English and who is not sure about the meaning of words such as "physical objects", "red" or "hand"; thus, for instance, someone who is learning English could ask: "Is this my wrist?" Similarly, the sentence "this is my hand" can be uttered to explain the meaning of the English word "hand" or to teach someone the use of the possessive in English (OC 369, and also McGinn 2022: 13-14):

"Only in certain cases is it possible to make an investigation "is that really a hand?" (or "my hand"). For "I doubt whether that is really my (or a) hand" makes no sense without some more precise determination. One cannot tell from these words alone whether any doubt at all is meant - nor what kind of doubt." (OC 372)

To sum up, I do hope I've been able to show that when Wittgenstein wrote the *Tractatus*, he thought that, thanks to sign-language, it was possible to resolve the question of regressum from the meaning of a sentence to the truth of a second sentence; however, once he rejected the notion of sign-language, he had to acknowledge the existence of "hinge propositions" concerning the meaning of other sentences.

References

- Baldwin, Tom (2010) "George Edward Moore", *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/sum2010/entries/moore/>](https://plato.stanford.edu/archives/sum2010/entries/moore/).
- Barba, Lorenzo (2023) "On the Nature of 'Names' in Wittgenstein's Tractatus", in: Alois Pichler / Esther Heinrich-Ramharter / Friedrich Stadler (eds.): *Contributions to International Wittgenstein Symposium 2023, 100 Years of Tractatus Logico-Philosophicus — 70 Years after Wittgenstein's Death. A Critical Assessment*. Online: <https://symposium.alws.at/#/article/4b4e5d15-4eb8-4132-8ebe-87abaaf9b77c>
- Coliva, Annalisa (2012) *Scetticismo: dubbio, paradosso e conoscenza*. Roma: Laterza.
- Coliva, Annalisa (2016) "Which Hinge Epistemology?" in: *International Journal for the Study of Skepticism* 6 (2-3), 79-96.
- Conant, James (1998) "Wittgenstein on meaning and use", *Philosophical Investigations* 21 (3), 222–250.
- Diamond, Cora (1991) *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*, Cambridge, MA, USA: MIT Press
- Frascolla, Pasquale (2000) *Il Tractatus logico-philosophicus di Wittgenstein*, Carocci, Roma.
- Gargani, Aldo Giorgio (1993) *Introduzione a Wittgenstein*. Roma-Bari: Laterza.
- Gargani, Aldo G. (2008) *Wittgenstein. Musica, Parola Gesto*, Milano: Raffaello Cortina Editore.
- McGinn, Marie (2022) *Wittgenstein, Scepticism and Naturalism: Essays on the Later Philosophy*. London: Anthem Press.
- Moore, George Edward (1942) 'A Reply to My Critics' in: P. A. Schilpp (ed.) *The Philosophy of G. E. Moore*, Northwestern University Press, Evanston ILL, 535-677.
- Ridpath, Ian (2018) "Plank Era" in Dictionary of Astronomy, Oxford University Press, Oxford. Online under: <https://www.oxfordreference.com/display/10.1093/acref/9780191851193.001.0001/acref-9780191851193-e-2880?rskey=uiToEC&result=1>
- Russell, Bertrand A.W. (1918) *Mysticism and Logic and other Essays*. London: Allen and Unwin.
- Wittgenstein, Ludwig (1961) "Notes dictated to G. E. Moore in Norway". In: Notebooks 1914–1916, S. 107–118. Hg. von G. H. von Wright und G. E. M. Anscombe, übersetzt von G. E. M. Anscombe. Oxford: Basil Blackwell. [Englisch] in https://www.wittgensteinproject.org/w/index.php?title=Notes_Dictated_to_G.E._Moore_in_Norway
- Wittgenstein, Ludwig (1929) "Some Remarks on Logical Form", *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 9, *Knowledge, Experience and Realism*, 162-171.

Wittgenstein, Ludwig (1968) *Philosophical Investigations*, Hg. von G. E. M. Anscombe und R. Rhees, übersetzt von G. E. M. Anscombe. Third edition with index. Oxford: Basil Blackwell. [Mit einem Register von Garth Hallett. Englische Übersetzung]

Wittgenstein, Ludwig (1969) *On Certainty / Über Gewißheit*, (OC) Hg. von G. E. M. Anscombe und G. H. von Wright, übersetzt von Denis Paul und G. E. M. Anscombe. Oxford: Basil Blackwell. [Deutsch mit englischer Übersetzung].

Wittgenstein, Ludwig (1963) TLP, *Tractatus Logico-Philosophicus*, Übersetzt von D. F. Pears und B. F. McGuinness. International Library of Philosophy and Scientific Method. Second impression, with a few corrections. London: Routledge and Kegan Paul.

Wittgenstein, Ludwig (1994) *Wiener Ausgabe Band 1. Philosophische Bemerkungen*, Hg. von Michael Nedo. Wien, New York: Springer.

Inside and Outside the Boundaries: Probability in Wittgenstein's *Tractatus*

Matteo Bizzarri (Pisa, Italy)

Abstract

The relationship between probability and classical logic can be approached from various angles. While the prevailing perspective often views probability as an extension of classical logic, there exists a less conventional approach that involves interpreting probability within the framework of classical logic itself. This alternative viewpoint, though less common, holds considerable interest and is exemplified in the works of philosophers such as Wittgenstein, De Finetti, Makinson, among others. Here, our focus lies on Wittgenstein's contribution, which holds both historical and philosophical significance in bridging probability and classical logic.

In his *Tractatus*, Wittgenstein introduced a method for computing probability using truth tables, which subsequently influenced the work of scholars like Carnap and Ramsey. Despite its historical importance, Wittgenstein's method has often been overlooked in the literature. Some scholars have interpreted it as an extension of the indifference principle, while others have seen it as an exploration of the relationship between beliefs and logic. Wittgenstein's method involves comparing two propositions: one analyzed solely in instances of truth, while the other is considered only when the first holds true. Remarkably, this approach bears resemblance to Makinson's supraclassical logic, albeit with differing methodologies.

This study aims to clarify Wittgenstein's method and its connection to probability and classical logic, with a particular focus on resolving the Lottery Paradox within the framework established by Wittgenstein.

1. Introduction

In a separate publication (Bizzarri 2024), we endeavored to offer a thorough analysis of Wittgenstein's concept of probability, demonstrating how it resolves or, better, dissolves the lottery paradox within his framework. Here, we aim to delve into the philosophical aspect of the coherence between probability and classical logic, starting from probability as presented in Wittgenstein's *Tractatus*.

Wittgenstein defines probability in terms of the relationship between "belief's truth-possibilities" (*Wahrheitsmöglichkeiten*) (Figueiredo 2023; Hay 2022; Cuffaro 2010; Ongaro 2021) and the truth possibilities of the proposition under consideration. Throughout his *Tractatus* (Wittgenstein 1922), Wittgenstein asserts that probability is inherently a priori, a stance he maintains in his later works where he vehemently rejects *frequentism* as the correct interpretation of probability.

Let's assume that someone playing dice every day were to throw, say, nothing but ones for a whole week, and that he does this with dice that turn out to be good when subjected to all other methods of testing, and that also produce the normal results when someone else throws them. Does he now have reason to assume a natural law here, according to which he always has to throw ones? Does he have reason to believe that things will continue in this way – or (rather) to assume that this regularity won't last much longer? So does he have reason to quit the game since it has turned out that he can throw only ones; or to continue playing, because now it is just all the more likely that on the next try he'll throw a higher number? – In actual fact he'll refuse to acknowledge the regularity as a law of nature; at least it will have to last for a long time before he'll consider this view of regularity. But why? – I think it's because so much of his previous experience in life refutes such a law, experience that has to be, so to speak – vanquished before we accept a totally new way of looking at things. (Wittgenstein BT, §33.7)

The philosophical idea of this paper lies on the fact that for Wittgenstein probability is a sort of extension of classical logic:

It is in this way that probability is a generalization.

It involves a general description of a propositional form.

We use probability only in default of certainty—if our knowledge of a fact is not indeed complete, but we do know something about its form.

(A proposition may well be an incomplete picture of a certain situation, but it is always a complete picture of something.)

A probability proposition is a sort of excerpt from other propositions. (TLP 1922, §5.156).

In Wittgenstein's conception of probability, truth is not solely dictated by logic but also by knowledge, specifically beliefs. Consequently, propositions that don't conform to classical logic as tautologies can still be ascribed non-zero values within a probabilistic framework.

Furthermore, we establish a correlation between the well-known Lottery Paradox and Wittgenstein's concept of probability, showcasing its straightforward resolution within this framework while also presenting intriguing

philosophical implications. By delving into these interconnections, our objective is to elucidate the distinctive characteristics and ramifications of Wittgenstein's probabilistic approach.

In the following two sections, we will revisit the concepts introduced in (Bizzarri 2024), omitting the details covered in the referenced paper. Additionally, in the third paragraph, we will present a philosophical argument that stems from probability in Wittgenstein's *Tractatus* and extends to the relationship between probability and Classical Logic.

2. Probability in the *Tractatus*

Wittgenstein's early notions regarding probability were first deliberated within the Circle of Vienna (Wright 1969) before undergoing refinement and solidification in the *Tractatus Logico-Philosophicus*. The treatment of probability in the *Tractatus* might seem peculiar at first glance, especially when contrasted with the conventional contemporary understanding of probability. Wittgenstein's distinct characterization of probability is elucidated in proposition 5.15:

If Tr is the number of the truth-grounds of a proposition r , and if Trs is the number of the truth-grounds of a proposition s that are at the same time truth-grounds of r , then we call the ratio Trs/Tr the degree of probability that the proposition r gives to the proposition s (TLP 1922, §5.15).

To understand better let's consider an example:

Example 2.1. Now, let's explore a common example from everyday life: flipping a coin. The central proposition we'll focus on is denoted as $x \underline{\vee} y$, where $\underline{\vee}$ signifies the mutually exclusive disjunction. In this scenario, the two potential outcomes, "heads" and "tails," are mutually exclusive. The truth table for the proposition $x \underline{\vee} y$ is as follows:

	$x \underline{\vee} y$	x	y
1	F	T	T
2	T	T	F
3	T	F	T
4	F	F	F

When considering only the instances where $x \vee y$ holds true, we observe that only the second and third rows meet this criterion. Let's now calculate the probabilities of x and y given the proposition $x \vee y$. For proposition x , among the two instances where $x \vee y$ is true, only the second instance has x as true, while the third instance has x as false. Consequently, the probability of x given $x \vee y$ is $1/2$. Similarly, for proposition y , among the two instances where $x \vee y$ is true, only the third instance has y as true, whereas the second instance has y as false. Thus, the probability of y given $x \vee y$ is also $1/2$.

In summary, when flipping a coin and considering the mutually exclusive disjunction proposition $x \vee y$, the probabilities of x and y given this proposition are both $1/2$, as anticipated.

3. Kolmogorov's axioms and Wittgenstein truth tables

Wittgenstein's truth tables satisfy Kolmogorov's axioms, validated in the *Tractatus*. The axioms, informally established in previous work, are:

- (K1) $0 \leq p(x) \leq 1$
- (K2) $p(x) = 1$ for some formula x
- (K3) $p(x) \leq p(y)$ whenever $x \vdash y$
- (K4) $p(x \vee y) = p(x) + p(y)$ whenever $x \vdash \neg y$

(K1) and (K2) derive from construction, bounded between 0 and 1. (K3) is validated via a truth table, substituting $x \vdash y$ with $x \rightarrow y$ as true. To indicate the probability of a certain proposition x given y we write $p_y(x)$.

<i>K3</i>	$x \rightarrow y$	x	y
1	T	T	T
2	<i>F</i>	<i>T</i>	<i>F</i>
3	T	<i>F</i>	T
4	T	<i>F</i>	<i>F</i>

where $p_{x \rightarrow y}(x) = 1/3$ and $p_{x \rightarrow y}(y) = 2/3$, so $p_{x \rightarrow y}(x) \leq p_{x \rightarrow y}(y)$ and (K4) can be proved by the following:

<i>K4</i>	$x \rightarrow \neg y$	x	y	$x \vee y$
1	<i>F</i>	<i>T</i>	<i>T</i>	<i>T</i>
2	T	T	<i>F</i>	T
3	T	<i>F</i>	T	T
4	T	<i>F</i>	<i>F</i>	<i>F</i>

where $p_{x \rightarrow y}(x) = 1/3$, $p_{x \rightarrow y}(y) = 1/3$ and $p_{x \rightarrow y}(x \vee y) = p_{x \rightarrow y}(x) + p_{x \rightarrow y}(y) = 1/3 + 1/3 = 2/3$ as wanted.

If we want to prove something generic the things become a little bit worse, because we have to check every case, for example if we want to prove (K5) $p(\neg x) = 1 - p(x)$ we must distinguish between the four combination of truthfulness and falsehood.

<i>K5</i>	<i>Formula</i>	x	$\neg x$
1	T	T	<i>F</i>
2	<i>F</i>	<i>F</i>	<i>T</i>

$p_{\text{formula}}(x) = 1$, $p_{\text{formula}}(\neg x) = 0$ and $p_{\text{formula}}(\neg x) = 1 - p_{\text{formula}}(x)$.

<i>K5</i>	<i>Formula</i>	x	$\neg x$
1	T	T	<i>F</i>
2	T	<i>F</i>	T

$p_{\text{formula}}(x) = 0.5$, $p_{\text{formula}}(\neg x) = 0.5$ and $p_{\text{formula}}(\neg x) = 1 - p_{\text{formula}}(x)$.

<i>K5</i>	<i>Formula</i>	x	$\neg x$
1	<i>F</i>	<i>T</i>	<i>F</i>
2	T	<i>F</i>	T

$p_{\text{formula}}(x) = 0$, $p_{\text{formula}}(\neg x) = 1$ and $p_{\text{formula}}(\neg x) = 1 - p_{\text{formula}}(x)$.

<i>K5</i>	<i>Formula</i>	x	$\neg x$
1	<i>F</i>	<i>T</i>	<i>F</i>
2	<i>F</i>	<i>F</i>	<i>T</i>

This last case is obviously special because we are giving a contradiction formula as a belief, so it's always false. Despite this, it was not really useful proving K5 from a formal point of view, because once K1-K4 were proved, than also K5 is provable from the first four axioms without using the truth tables.

Proving the Kolmogorov's axioms has a double benefit: it proves that Wittgenstein's idea of probability is something related to the common idea of it and it permits us to restrict the set of valuations to make a supraclassical logic.

4. Generalization of Wittgenstein's probability

Wittgenstein's probability offers a consistent probabilistic logic within classical limits, enabling resolution of belief paradoxes like the Lottery Paradox. By extending classical logic, we maintain conjunction principles, contrary to previous suggestions. This approach addresses paradoxes effectively, notably the Lottery Paradox (Hawthorne 2009; Foley 1992; Leitgeb 2017; Kyburg 1961), which persists under classical frameworks. The Lottery Paradox can be formulated as follows:

"Let's consider a fair 1000-ticket lottery that has only one winning ticket. A perfectly rational agent knows that each ticket has a probability of 999/1000 of not winning. Thus, it is rational for the agent to accept that each ticket will not win because this probability is greater than her Lockean threshold. This reasoning can be extended to every other ticket in the lottery, leading to the conclusion that somehow every ticket will not be the winning ticket. However, the lottery is fair, so the conjunction of all these statements has to be false, rather than true as it appears."

The idea of solving this paradox thanks to Wittgenstein's idea is interesting because of the proposition 5.156 that we have addressed in the introduction. Leveraging Wittgenstein's notion of probability as a generalization, we demonstrate a method to resolve the paradox. This involves establishing a unique True line amidst a conjunction of numerous negative propositions, maintaining its position as propositions vary. Utilizing this insight, we construct a disjunction to encompass all scenarios, yielding exactly n True lines, where n is the count of literals within the formula.

	$\neg p_1$	\wedge	$\neg p_2$	\wedge	\dots	\wedge	p_x	\wedge	$\neg p_{x+1}$	\wedge	\dots	\wedge	$\neg p_n$
1	F		F		...		T		F		...		F
2	F		F				T		F				T
\vdots	\vdots		\vdots				\vdots		\vdots				\vdots
2^{n-1}	F		T				T		T				T
$2^{n-1} + 1$	T		F				F		F				F
\vdots	\vdots		\vdots				\vdots		\vdots				\vdots
$2^{n-1} + 2^{n-2}$	T		F				T		F				T
$2^{n-1} + 2^{n-2} + 1$	T		T				F		T				F
\vdots	\vdots		\vdots				\vdots		\vdots				\vdots
$2^n - 2^{n-x} - 1$	T		T				T		T				F
$2^n - 2^{n-x}$	T		T				T		T				T
$2^n - 2^{n-x} + 1$	T		T				F		F				F
\vdots	\vdots		\vdots				\vdots		\vdots				\vdots
2^n	T		T				F		F				T

This truth table needs some hint to let it be cleared:

- Highlighting significant transitions enhances clarity. For instance, 2^{n-1} marks the last row where $\neg p_1$ changes, at the midpoint of the table. Similarly, $2^{n-1} + 2^{n-2}$ precedes the change of $\neg p_2$.
- The row $2^n - 2^{n-x}$ is notable, filled entirely with T. It results from doubling T instances left of p_x while halving them right of p_x , ending with a single T for $\neg p_n$.
- $2^n - 2^{n-x}$ signifies the last row before p_x changes, equivalent to summing halved values successively, reflecting diminishing T instances.

The following theorem is the main theorem to be proved in order to generalize Wittgenstein's probability proved in (Bizzarri 2024):

Theorem 3.1 If a proposition made by an arbitrary number of elementary letters is made by all negated formulas and one positive formula, the only line

that is made by true instances is the line marked with the number $2^n - 2^{n-x}$, where x is the position of the elementary letter starting from the left.

Thanks to Theorem 3.1, we can observe how the truth table regarding the Lottery Paradox can be resolved. Interestingly, a similar outcome was presented in Bizzarri 2024, albeit through a completely different method. However, we will delve into this in the next section. Surprisingly, the Lottery Paradox can also be solved in Fractional Semantics as presented in (Bizzarri 2023) and also in the limits of Classical Logic.

5. Probability and classical logic

Probability has long presented a challenging relationship with classical logic. On one hand, it appears to extend classical logic's, yet on the other hand, it seems to impose constraints on its rules. Within Wittgenstein's conceptual framework, probability finds a place within the limits of classical logic, particularly in his exploration of the interplay between beliefs and propositions.

By the way Wittgenstein's method was seen as a generalization over possibilities, also if the question is more challenging. In De Finetti's "Theories of probabilities" (Finetti 1931), a clear distinction is drawn between possibilities, which are objective, and probabilities, which are subjective. Wittgenstein's perspective occupies a middle ground between these concepts. While he meticulously analyzes each possibility of falsity and truthfulness akin to De Finetti's framework, the subjective element emerges from the agent's capacity to select the initial set of propositions, intertwined with the agent's personal comprehension of a given argument.

Despite its significance, probability in Wittgenstein's oeuvre is often treated as peripheral, with scant exploration of his specific viewpoints on the subject. Notably, Wittgenstein's fundamental musings on the nature of probability are encapsulated in the *Tractatus*, commencing from proposition 5.1:

Truth-functions can be arranged in series. That is the foundation of the theory of probability (TLP 1922, §5.1).

In essence, Wittgenstein's exploration of probability can be interpreted as a compelling endeavor to bridge the gap between beliefs and propositions, but also as the first tentative of include probability into the limits and confines of Classical Logic. While it shares similarities with objective possibilities, it also exhibits subjective features by empowering agents to shape the initial set of propositions according to their individual knowledge. Despite its relatively limited exposition, Wittgenstein's reflections on probability in the *Tractatus* offer invaluable insights into this intricate domain. Initial reflections on these can be traced back to the Notebooks 1914-1916 and discussions within the Vienna Circle.

In this argument, I posit that Wittgenstein's notion of probability, despite facing substantial critique—many of which have been aptly addressed by Cuffaro in (Cuffaro 2010)—remains a significant exemplar of the symbiotic relationship between classical logic and probability theory. This synergy has been further advanced by eminent philosophers such as Ramsey and De Finetti, who have embraced and expanded upon this conceptual interplay in their respective works. For instance, De Finetti elucidates in "Theories of Probabilities" that probability inherently resides within subjective realms, encapsulating one's "degree of beliefs." This intrinsic link between classical logic and the subjective assessment of probabilities is widely acknowledged and appreciated within philosophical discourse.

Moreover, what Wittgenstein suggested in the *Tractatus*, i.e., that probability is a relationship between beliefs and the logic, will be a firm point also in his later writings. For example in the Big Typescript he writes:

Induction is a process based on an economic principle. (BT §33.3)

Articulating a notion that resonated strongly with De Finetti, it becomes evident that probability is inherently grounded in subjective interpretation and operates on an economic principle. By leveraging beliefs alongside classical logic, the framework fundamentally aligns itself with Classical Logic, thus situating probability within the confines of Classical Logic and concurrently diminishing its boundaries due to Post-Completeness. This perennial issue arises when beliefs are contextualized within Classical Logic, necessitating a trade-off between consistency and structural integrity. The

forfeiture of structurality precludes the utilization of Substitution, one of the fundamental operation within classical logic.

Concluding, Wittgenstein's view on probability has several peculiar aspects. If we follow Wittgenstein's idea until the very end, his view on probability remains within the boundaries of Classical Logic (it is, in fact, only a generalization, but the structure remained the same), and, as we have shown, it also satisfies Kolmogorov's axioms and resolves the Lottery Paradox. These significant aspects aid in understanding how probability can be constructed within or outside Classical Logic. Expanding the boundaries of Classical Logic is technically challenging but straightforward: it suffices to add semantics that can reconnect our logic to the mathematical form of probability. Conversely, staying within the boundaries of Classical Logic is more difficult to justify but technically simpler and philosophically more intriguing. We believe that Wittgenstein was able to grasp many of the problems that logicians still face today when dealing with Probability and Classical Logic, and he resolved them in an elegant and synthetic manner. We propose that this initial attempt served as the foundation upon which De Finetti and Ramsey based their work, and its philosophical significance must be revitalized.

6. Conclusions

In this paper we provided a description of the first Wittgenstein's view on probability.

In our paper, we have extensively tackled the challenges posed by Wittgenstein's probabilistic framework, particularly focusing on the Lottery Paradox. At first glance, Wittgenstein's approach to probability may seem unorthodox, but upon closer examination, it reveals a coherent structure that aligns with Kolmogorov's axioms and qualifies as a supraclassical logic.

Our research underscores the consistency of Wittgenstein's perspective, offering a resolution to the Lottery Paradox within this framework. What was once considered a paradox now finds clarity through an extension of classical logic.

While our methodology isn't a complete departure from conventional approaches, it deserves more attention for its innovative incorporation of beliefs into the analysis of probability. This inclusion adds a fresh dimension to the field and sets the stage for the development of a robust supraclassical probabilistic logic.

Looking forward, we anticipate that our exploration of supraclassical logic and probabilistic reasoning, enriched by Wittgenstein's philosophical insights, will contribute significantly to the establishment of a solid foundation bridging logic and philosophy.

References

- Bizzarri, Matteo (2023). "A solution to the Lottery Paradox through Fractional Semantics", in: *Fifth Pisa Colloquium in Logic, Language and Epistemology*. Ed. by Giacomo Turbanti Luca Bellotti Enrico Moriconi. Edizioni ETS, 15–29.
- Bizzarri, Matteo (2024). "Wittgenstein, probability and supra-classical logics", forthcoming.
- Cuffaro, Michael E. (2010) "Wittgenstein on Prior Probabilities", in: *Proceedings of the Canadian Society for History and Philosophy of Mathematics* 23, 85–98.
- De Finetti, Bruno (1931), *La logica dell'incerto*, Libreria Editrice Francesco Perrella.
- De Finetti, Bruno De (1989), *La logica dell'incerto*, Il Saggiatore.
- Figueiredo, Florian Franken (2023) "Wittgenstein's 1929–30 inquiries into probability", *Philosophical Investigations* 46.1, 50–75.
- Foley, Richard (1992) "The Epistemology of Belief and the Epistemology of Degrees of Belief", *American Philosophical Quarterly* 29.2, 111–124.
- Hajek, Alan (2019) "Interpretations of Probability", in: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2019. Metaphysics Research Lab, Stanford University.
- Halpern, Joseph Y. (2003) *Reasoning about Uncertainty*, Cambridge, MA, USA: MIT Press.
- Hawthorne, James (2009) "The Lockean Thesis and the Logic of Belief", in: *Degrees of Belief*. Ed. by Franz Huber and Christoph Schmidt-Petri. Synthese Library: Springer, 49–74.
- Hay, Clare (2022) "Probability in Wittgenstein's *Tractatus*", *Philosophical Investigations* 45.2, 130–145.
- Kyburg, Henry Ely (1961) *Probability and the Logic of Rational Belief*, Middletown, CT, USA: Wesleyan University Press.

Leitgeb, Hannes (2017) *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press.

Makinson, D. (2005) *Bridges from classical to nonmonotonic logic*. Lightning Source, Milton Keynes.

Ongaro, Malvina (2021) "The Interpretation of Probability in the *Tractatus Logico-Philosophicus*", *Wittgenstein-Studien* 12.1, 131–144.

Paris, J. B. (1995) *The Uncertain Reasoner's Companion: A Mathematical Perspective*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Piazza, M. and G. Pulcini (Oct. 2016) "Uniqueness of axiomatic extensions of cut-free classical propositional logic", in: *Logic Journal of the IGPL* 24.5, 708–718.

Post, Emil (1990) "Introduction to a general theory of elementary propositions", in: *From Frege to Gödel: A Source Book in Mathematical Logic*, Ed. by Jean van Heijenoort. Harvard University Press.

Wittgenstein, Ludwig (1922) *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul.

Wittgenstein, Ludwig (1971) *Prototractatus an Early Version of Tractatus Logico-Philosophicus*. Edited by B.F. McGuinness, T. Nyberg and G.H. Von Wright, with a Translation by D.F. Pears and B.F. McGuinness. An Historical Introd. By G.H. Von Wright and a Facsim. Of the Author's Manuscript. Cornell University Press.

Wittgenstein, Ludwig (2012) *The Big Typescript: TS 213*, German English Scholars' Edition. Blackwell Pub; Bilingual.

Wright, G. H. Von (1969) "Wittgenstein's views on probability", *Revue Internationale de Philosophie* 23.88/89 (2/3), 259–279.

Toward a Dissolution of the Color-Exclusion “Problem”

Edgar Boedeker (Cedar Falls, USA)

Abstract

In the *Tractatus*, Wittgenstein makes numerous demands, including the following: (1) All propositions are sentential or, for the purposes of this paper, quantificational truth-functions of logically independent elementary propositions. (2) All necessary, impossible, or contingent propositions are logically true (tautologous), logically false (contradictory), or “senseful” (*sinnvoll*), respectively. (3) It’s “impossible” “for two colors to be in one place in the visual field at the same time” (6.3751); attributions of just one color to a place and time are contingent. Upon his public return to philosophy, Wittgenstein argued in “Some Remarks on Logical Form” that (1)-(3) are inconsistent, and (1) must be abandoned. What’s known in the literature as the “color-exclusion problem” is the puzzle of how to reconcile these demands. Employing quantification, three dyadic predicates, and the additive “RGB” color-model, I show how to accomplish this for attributions to visible objects of white, black, and pure primary and secondary colors. These analyses are “ontologically” neutral, thus bypassing longstanding debates concerning whether the *Tractatus* is committed to phenomenalism or physicalism. The syntax of the notation I propose mirrors the “logical form” of visible objects: all such objects have some color. I conclude by arguing this approach should be regarded not as providing a *solution* to a *problem*, but as *dissolving* a *pseudo-problem*.

In the *Tractatus*, Wittgenstein makes numerous demands, including the following:

Dem1: All propositions are sentential or, for the purposes of this paper, quantificational truth-functions of logically independent (4.211, 5.134, 5.152) elementary propositions (5, 5.3, 6-6.001).

Dem2: All necessary, impossible, or contingent propositions are logically true (tautologous), logically false (contradictory), or “senseful” (*sinnvoll*), respectively (5.525; cf. 6.375).

Dem3: It’s “impossible” “for two colors to be in one place in the visual field at the same time” (6.3751). Attributions of just one color to a place and time are contingent.

Consider the sentence “Just red is in place P at time T and just blue is in P at T ”, expressed as “ $RPT \wedge BPT$ ” (RLF 1929: 168-171). If “ RPT ” and “ BPT ” are elementary propositions, Dem1 implies it can be expressed in a standard, four-row truth-table (figure 1).

<i>RPT</i>	<i>BPT</i>	
T	T	T
T	F	F
F	T	F
F	F	F

Figure 1.

Dem2-Dem3 imply this proposition is a “contradiction” (6.3751). The four-row truth-table, however, wrongly displays it as if it were senseful, hence contingent, as seen in its first row. Its proper expression (figure 2) eliminates this row (RLF 1929: 169f).

<i>RPT</i>	<i>BPT</i>	
T	F	F
F	T	F
F	F	F

Figure 2.

But this violates Dem1. Dem1-Dem3 thus appear inconsistent.

At 6.3751, Wittgenstein proposes dealing with this difficulty as follows:

Consider how this contradiction manifests itself in physics – roughly like this: that a particle cannot have two velocities at the same time; that is, that it cannot be in two places at the same time; i.e., that particles in different places at one time cannot be identical.

Frank Ramsey argued this “solution” is untenable, for “Wittgenstein is only reducing the difficulty to that of the *necessary* properties of space, time, and matter”; but these “properties [...] are hardly capable of a further reduction” (1923: 31) to logical truths or falsehoods. If no such analyses can be found, Dem1-Dem3 remain inconsistent.

This puzzle is known in the literature as the “color-exclusion problem” or “color-incompatibility problem”. Upon his public return to philosophy, Wittgenstein (RLF 1929: 168-171) maintained it’s insoluble, inferring he must abandon Dem1. This is often regarded as the first crack in the edifice of the *Tractatus* (e.g., Jacquette 1997: 153-192, Medina: 2002: 5-53). Moss (2012) and Hintikka and Hintikka (1986: 121-124) have proposed “solutions”, but they face

difficulties; cf. Welty (2021) on Moss, and Suárez (2014: 36-40) on Hintikka and Hintikka.

Here, I’ll propose a way to analyze certain color-attribution statements that satisfies Dem1-Dem3, and Dem4-Dem5 below. Employing quantification, three dyadic predicates, and the “RGB” color-model, it allows for the expression of attributions to visible objects of white, black, and pure primary and secondary colors. I conclude by arguing this approach should be regarded not as providing a *solution* to a *problem*, but as *dissolving* a *pseudo-problem*.

Canfield (1976; cf. Chandler [2023]) proposes a way to satisfy Dem1-Dem3. It will prove helpful to present my proposal through an explication and critique of his.

He analyzes color-attribution sentences as sentential truth-functions of elementary propositions formed using Wittgenstein’s triadic predicate xPT . The values of x are degrees of primary colors red, blue, green, and yellow, referred to by names “ r_1 ”, ..., “ r_j ”, “ b_1 ”, ..., “ b_k ”, “ g_1 ”, ..., “ g_m ”, “ y_1 ”, ..., “ y_n ”. “ RPT ”, e.g., gets analyzed as $RPT1$:

$$\begin{aligned} & “r_1PT \wedge \neg r_2PT \wedge \dots \wedge \neg r_jPT \wedge \neg b_1PT \wedge \neg b_2PT \wedge \dots \wedge \neg b_kPT \wedge \\ & \neg g_1PT \wedge \neg g_2PT \wedge \dots \wedge \neg g_mPT \wedge \neg y_1PT \wedge \neg y_2PT \dots \wedge \neg y_nPT”, \end{aligned}$$

the ellipses standing for the conjunctions of the negations of the appropriate remaining elementary propositions. “ BPT ” is analyzed as $BPT1$, resulting from $RPT1$ by replacing “ r_1PT ” with “ $\neg r_1PT$ ”, and “ $\neg b_1PT$ ” with “ b_1PT ”.

Canfield treats non-primary colors as mixtures of degrees of primary ones (cf. BT 2005: 340f). Here, he employs a “subtractive” color-model: although degrees of primary colors are *not* pigments (BT 2005: 341, PR 1975: 273), they possess combinatory properties mirroring them. “Orange is in P at T ” (“ OPT ”), e.g., is analyzed as $OPT1$, resulting from $RPT1$ by replacing “ $\neg y_1PT$ ” with “ y_1PT ”.

Since Canfield’s approach allows for the possibility that degrees of different primary colors occupy the same spatiotemporal location, it satisfies Dem1.

His analyses also form the basis of a way to satisfy Dem2-Dem3. E.g., analyzing " $RPT \wedge BPT$ " as the conjunction of $RPT1$ and $BPT1$ shows it's contradictory, hence impossible. Accordingly, necessary truth

$$"RPT \rightarrow \neg BPT"$$

is manifestly tautologous, and " RPT " and " BPT " senseful, i.e., contingent.

Wittgenstein, however, makes a further demand:

Dem4: "There is one and only one complete analysis of a proposition" (3.25).

The propositions to be analyzed here are such color-attributions as " RPT ". There's a one-to-one relation between these propositions and what I'll call possible (truthmaking) "visible situations": the presence of just one color in P at T . In four ways, CD1-CD4, Canfield's proposal fails to satisfy Dem4. CD1-CD3 concern its yielding more than one analysis of certain color-attribution sentences, something Wittgenstein deems "obviously absurd" (1929: 168). Every proposition, after all, has just one set of truth-conditions. Indeed, since all Canfield-style analyses are mutually contradictory, his proposal entails there would be contingent color-attribution sentences with mutually contradictory analyses.

CD1, endemic to subtractive color-models, pertains to attributions of just one primary color to a spatiotemporal location. Since a single "dab" of a color in P at T looks the same as two or more "dabs" of it, $RPT1$, e.g., describes the same visible situation (RPT) as does $RPT2$, resulting from $RPT1$ by replacing " $\neg r_2PT$ " with " r_2PT ".

CD2 involves attributions of mixtures of primary colors. Since degrees of the same color are indistinguishable, analyses differing only by interchanging their names describe the same visible situation. $OPT1$, e.g., describes the same one (OPT) as does $OPT2$, resulting from $OPT1$ by replacing " r_1RT " with " $\neg r_1RT$ ", " $\neg r_2RT$ " with " r_2RT ", " y_1PT " with " $\neg y_1PT$ ", and " $\neg y_2PT$ " with " y_2PT ".

CD3 concerns the joint negation of each attribution of a degree of primary color. It becomes pressing when we complete the color-octahedron by

supplementing Canfield’s primary colors with black and white (PR 1975: 51f, 278f; WWK 1984: 42f), required to attribute not just hue, but also brightness. Since it makes no sense to try to say that a “patch” (2.0131) with positive extension in two dimensions (RLF 1929: 166) is colorless, i.e., invisible, this sentence is either nonsensical or attributes the color of the background “canvas” (cf. PR 1975: 115) to this spatiotemporal location. But the latter proposition asserts the existence of the same visible situation as does the proposition explicitly attributing this color to P at T . Here again, we’d absurdly have mutually contradictory analyses of the same color-attribution.

CD4 lies in Canfield’s analyses failing to be “complete” (RLF 1929: 167; cf. PR 1975: 108f), i.e., to specify *exactly* what color is said to be at a spatiotemporal location. For they only state that the *explicitly named* degrees of primary colors are or aren’t in P at T . It’s thus unclear whether Canfield’s analyses lay out the full truth-conditions of the propositions they purport to analyze. To eliminate this “indetermina[cy]” (MS 105: 82), we’d have to conjoin to his analyses a “completing supplementary statement” (RLF 1929: 169) asserting that no other colors are there and then. Wittgenstein suggests such an “addition” can’t “be made” (PR 1975: 109). In standard predicate logic,

$$\neg(\exists x)(xPT \wedge p)$$

might fit the bill, where p is the conjunction of all expressions $\neg x = z$ for every degree-of-a-primary-color z attributed to P at T . But this departs from Canfield’s non-quantificational analyses, opening the door to the quantificational ones I’ll propose.

To satisfy Wittgenstein’s demands, I’ll adopt not a subtractive color-model, but an “additive” one. *In the case of* physical colors, the one I’ll be employing works by bathing an otherwise unilluminated black surface in combinations of red, green, and blue colored lights of various intensities. It’s well known that colors of (virtually) any hue, saturation, and brightness can be determined with this “RGB” color-model. I’ll use the “classical” model, not one employed in graphics applications. The color space of my analyses is discrete, not continuous. This tallies with Wittgenstein’s claim that all “systems for

describing the world” have “a *specific* fineness of grain”, such that we “can always get as close as [we] want to” the world’s being “*completely* described” (6.341-6.342).

Relations among electromagnetic light, color as perceived, and various color spaces are highly complex. Here, I’m *not* proposing a physical, psychological, or physiological *theory* of color or color-perception, but a way to analyze *logically* certain color-attribution statements. The sole purpose of the analyses I provide is to satisfy Dem1-Dem5, and that’s just how they should be understood. They’re “ontologically” neutral as to whether visible objects occur in a “phenomenological” sphere of sense-data (a view Wittgenstein maintained in most of 1929) or in the physical world, thus appropriately circumventing longstanding debates concerning whether the *Tractatus* – a work not in “epistemology” (4.1121), but the philosophy of *logic* – is committed to phenomenalism (e.g., Hintikka and Hintikka [1989: 137-175]) or physicalism (e.g., Lampert [2000]).

Here I’ll bracket spatiotemporal locations, focusing exclusively on attributing colors to simple *objects* in the visual field. The analyses employ just three dyadic predicates, which I’ll *elucidate* – *not* define – as follows:

$R^{vi}x^vy^i$: (visible object) x^v is-bathed-in-a-degree-of-red-light-by (illuminator) y^i .

$G^{vi}x^vy^i$: (visible object) x^v is-bathed-in-a-degree-of-green-light-by (illuminator) y^i .

$B^{vi}x^vy^i$: (visible object) x^v is-bathed-in-a-degree-of-blue-light-by (illuminator) y^i .

The ordered superscripts indicate which arguments may and may not occur in the first and second places of the predicates; thus that all and only visible objects a^v , b^v , etc., can be-bathed-in-a-degree-of-colored-light, and all and only “illuminators” a^i , b^i , etc., can thus bathe them.

These predicates therefore satisfy a final Tractarian demand, not met by Canfield’s analyses:

Dem5: Something is a visible object only if it can occur in a possible state-of-things (*Sachverhalt*) in which it has a color (2.013-2.0131); and it must be possible to construct a "correct concept-script" (5.534) whose "logical syntax" (cf. 3.325) precludes the formation of "nonsensical pseudo-propositions" (4.1272) not conforming to this "logical form" (2.0233; cf. 2.0251) of visible objects.

Whereas our predicates are "real functions" (4.126) corresponding to "real (external) relations" (4.122) among objects, "visible object" and "illuminator" express "formal concepts" (4.122, 4.126, 4.12721) corresponding to "formal", "internal", properties of objects: what's "unthinkable that [they don't] possess" (4.123). We employ predicates to say how visible objects and illuminators are related, but we can't say what formal properties they have; this "shows itself", via superscripts, "in the very sign[s] for the object[s]" (4.126; cf. 4.124).

Colors of visible objects are determined as follows. Let n be the total number of illuminators, the set of which is finite and non-empty. If visible object a^v is-bathed-in-a-degree-of-red-light-by r illuminators, is-bathed-in-a-degree-of-green-light-by g illuminators, and is-bathed-in-a-degree-of-blue-light-by b illuminators; then a^v has the color yielded by combining red, green, and blue lights of intensities

$$(r / n), (g / n), \text{ and } (b / n),$$

respectively, where 1 is full intensity.

Each illuminator can bathe any combination of visible objects independently in any combination of one or no degree of red, green, or blue light. Our predicates thus satisfy Dem1.

We can *imagine* things as follows. Illuminators are like white lights with equal intensity, each focused on every visible object via three conduits. One contains just a red filter, another green, the third blue. Each conduit is either closed (e.g., $\neg R^{vi}a^vc^i$) or open ("bathing", e.g., $R^{vi}a^vc^i$). With visible objects a^v and b^v and illuminators c^i and d^i , figure 3 illustrates

$$"R^{vi}a^vc^i \wedge G^{vi}a^vc^i \wedge \neg B^{vi}a^vc^i \wedge$$

$$\neg R^{vi}a^vd^i \wedge \neg G^{vi}a^vd^i \wedge \neg B^{vi}a^vd^i \wedge \\ \neg R^{vi}b^vc^i \wedge \neg G^{vi}b^vc^i \wedge B^{vi}b^vc^i \wedge \\ R^{vi}b^vd^i \wedge \neg G^{vi}b^vd^i \wedge \neg B^{vi}b^vd^i.$$

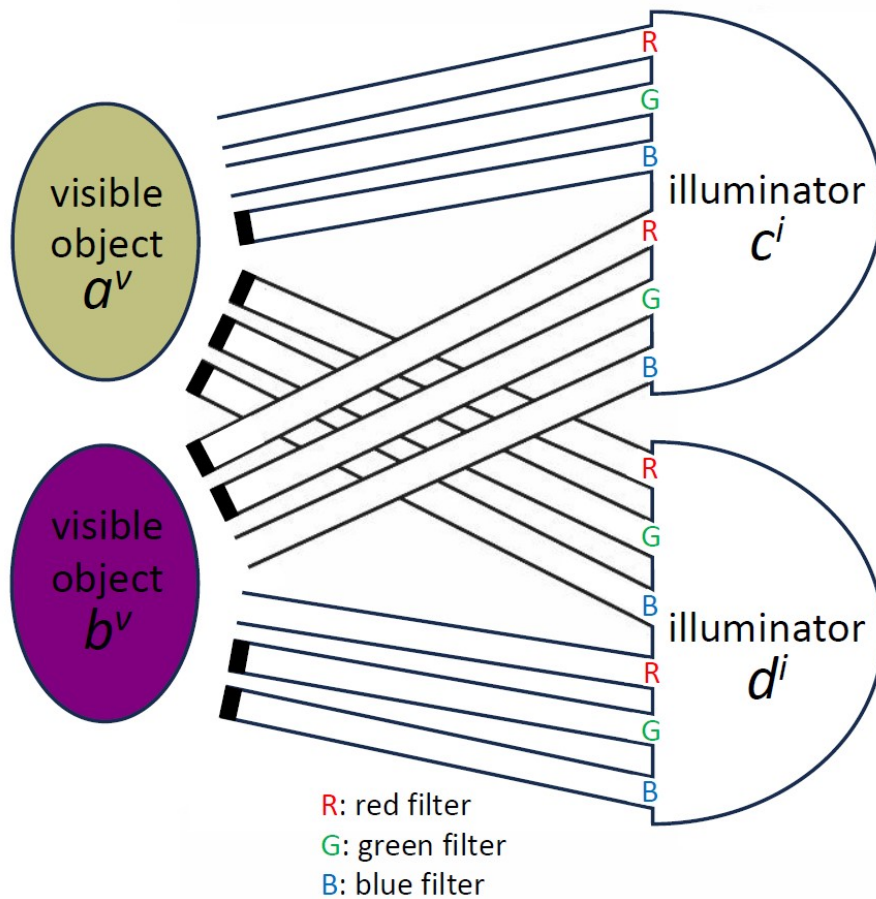


Figure 3.

Assuming c^i and d^i are the only illuminators, full-intensity colored light is two degrees. Since a^v is bathed (by c^i) just in one degree of red light and one of green, and b^v is bathed just in one degree of blue light (by c^i) and one of red (by d^i), a^v is dark yellow and b^v dark magenta.

Key to my approach is the fact that no constant names of illuminators occur in analyses of color-attribution sentences. Rather, each analysis contains bound variables – “ x^i ” in the analyses below – ranging over all illuminators.

I'll now provide some examples of such analyses, for perspicuity using standard notation instead of Wittgenstein's N-operator (5.5, 5.502, 6-6.001) and his way of eliminating the identity-sign (5.53-5.533):

(i) " a^v is just red":

$$"(\forall x^i)R^{vi}a^vx^i \wedge \neg(\exists x^i)G^{vi}a^vx^i \wedge \neg(\exists x^i)B^{vi}a^vx^i".$$

(ii) " a^v is just blue":

$$"\neg(\exists x^i)R^{vi}a^vx^i \wedge \neg(\exists x^i)G^{vi}a^vx^i \wedge (\forall x^i)B^{vi}a^vx^i".$$

(iii) " a^v is just [pure secondary color] magenta":

$$"(\forall x^i)R^{vi}a^vx^i \wedge \neg(\exists x^i)G^{vi}a^vx^i \wedge (\forall x^i)B^{vi}a^vx^i".$$

(iv) " a^v is just white":

$$"(\forall x^i)R^{vi}a^vx^i \wedge (\forall x^i)G^{vi}a^vx^i \wedge (\forall x^i)B^{vi}a^vx^i".$$

(v) " a^v is just black":

$$"\neg(\exists x^i)R^{vi}a^vx^i \wedge \neg(\exists x^i)G^{vi}a^vx^i \wedge \neg(\exists x^i)B^{vi}a^vx^i".$$

(vi) "Something is just black":

$$"(\exists y^v)(\neg(\exists x^i)(R^{vi}y^vx^i \wedge \neg(\exists x^i)G^{vi}y^vx^i \wedge \neg(\exists x^i)B^{vi}y^vx^i))".$$

Such analyses allow us to satisfy Dem2-Dem3. The analysis of " a^v is both just red and just blue", e.g., is manifestly logically false: the conjunction of the analyses of (i) and (ii), which I'll abbreviate as

$$(A-i) \wedge (A-ii).$$

Accordingly, analyses of such necessary truths as "If a^v is just red, then it's not just blue" are manifestly logically true; and analyses of attributions of just one color sensible, hence contingent.

My proposal also satisfies Dem4. I'll illustrate this by showing how, *mutatis mutandis*, it overcomes CD1-CD4; as far as I can see, these are the only difficulties with Canfield's approach. My use of an additive color-model avoids CD1; for whereas in subtractive models adding degrees of a single primary color doesn't change the visible situation described, adding degrees of a colored light does. This approach also circumvents CD2; although illuminators themselves are indistinguishable, since we attribute a color to a visible object via bound variables whose values are names of illuminators, just which instantiations make these quantificational statements true or false is

irrelevant. And since (v) is the analysis just of “ a^v is just black”, CD3 doesn’t arise. Finally, my proposal overcomes CD4; whereas we can’t tell from Canfield’s non-quantificational analyses whether there’s a degree-of-primary-color they don’t name, the use of bound variable x^i in our analyses allows us to make statements about *all* illuminators.

We can thus see that my approach satisfies Dem1-Dem5 for attributions of white, black, and the remaining pure primary color (green) and pure secondary colors (yellow and cyan).

It encounters a complication, however, for it can’t provide one and only one complete analysis of attributions of tertiary, quaternary, etc., colors. Take orange, resulting from combining just full-intensity red light with green light in a 2:1 ratio. a^v is orange *if*

$$\begin{aligned} & “(\exists x^i)(\exists y^i)(R^{vi}a^v x^i \wedge R^{vi}a^v y^i \wedge \neg x^i = y^i \wedge \neg(\exists z^i)(R^{vi}a^v z^i \wedge \neg z^i = x^i \wedge \neg z^i = y^i) \wedge \\ & (\exists u^i)(G^{vi}a^v u^i \wedge (u^i = x^i \vee u^i = y^i) \wedge \neg(\exists w^i)(G^{vi}a^v w^i \wedge \neg w^i = u^i)) \wedge \\ & (\forall x^i)R^{vi}a^v x^i \wedge \neg(\exists x^i)B^{vi}a^v x^i” \end{aligned}$$

is true. This analysis implies there are exactly two illuminators, so full-intensity colored light is two degrees. But a^v is orange also if the ratio is the same and the number of illuminators, hence degree of full-intensity colored light, is four, six, eight, etc. To yield the one-to-one relation, required by Dem4, between color-attributions and their analyses, we must analyze the proposition by asserting the “disjunction” of the members of the “formal series” (*Formenreihe*: 5.501, 4.1273) of propositions asserting that the 2:1 ratio obtains, each with a different even number of illuminators. In my (2023), I show how to construct propositions involving formal series.

I suggest we regard my approach to color-attribution statements not as providing a *solution* to a *problem* – a term generally used in the literature – but as *dissolving* a *pseudo-problem*. Here’s what I mean. Someone expresses p ’s being a *puzzle* for them when they ask a question expressible as “Why is ‘ p ’ true?”, “Why is ‘ p ’ impossible?”, etc. (I disregard cases where this isn’t the case for “ p ”.) Posing such a question opens the floor to two kinds of adequate replies. One is a true explanatory sentence expressible as “‘ p ’ is true (etc.) because q ”, where the proposition expressed by “ p ” occurs in the reply worded

only as in the question. Such a reply (partially) *solves* the puzzle. A puzzle is a *problem* if it can be solved. Electromagnetic theories of physical light and physiological theories of retinal functions provide (partial) solutions to the problem of why we perceive colors as we do.

A second kind of adequate reply *dissolves* the puzzle expressed in the question. Here, the reply formulates "*p*" in more perspicuous terms so it's no longer a puzzle for the person who had asked it. A puzzle is a *pseudo-problem* if it can be dissolved.

We can reformulate the puzzle raised by our initial example as "Why is 'a' is just red and just blue' impossible?", the proposition imperspicuously expressed as

$$"Ra^v \wedge Ba^v".$$

Our analyses permit the reply that it's impossible because it means

$$(A-i) \wedge (A-ii).$$

Since this is a manifest logical falsehood, the initial puzzle is dissolved, revealing it to be a pseudo-problem.

Here we have an example of the "activity" Wittgenstein at 4.122 posits as definitive of philosophy. Its "purpose is the logical clarification of thoughts" through "elucidations", i.e., analysis, the "result" of which "is not 'philosophical propositions', but propositions becoming clear." Such elucidations allow us to "avoid" (3.325) philosophical "confusions" (3.324) that "rest on a misunderstanding of the logic of our language" (Preface, ¶2; cf. 4.003). I hope to have accomplished this with puzzles concerning certain impossible color-attributions, and pointed a way toward doing so with others.

References

Boedeker, Edgar (2023) "Tractarian Cutlery: Formal Series and Equinumerosity", *Contributions of the Austrian Ludwig Wittgenstein Society* 29, 82-92.

Canfield, John V. (1976) "Tractatus Objects", *Philosophia* 6, 81-99.

Chandler, David (2023) "A Review of the Colour-Exclusion Problem at Tractatus 6.3751", *Contributions of the Austrian Ludwig Wittgenstein Society* 29, 104-110.

- Hintikka, Merrill; Hintikka, Jaakko (1989) *Investigating Wittgenstein*, New York: Blackwell.
- Jacquette, Dale (1997) *Wittgenstein's Thought in Transition*, West Lafayette: Purdue University Press.
- Lampert, Timm (2000) *Wittgensteins Physikalismus: Die Sinnesdatenanalyse des Tractatus Logico-Philosophicus in ihrem historischen Kontext*, Paderborn: Mentis.
- Medina, José (2002) *The Unity of Wittgenstein's Philosophy: Necessity, Intelligibility, and Normativity*, Albany: SUNY Press.
- Moss, Sarah (2012) "Solving the Color Incompatibility Problem", *Journal of Philosophical Logic* 41, 841-851.
- Ramsey, Frank (1923) "Critical Notice of L. Wittgenstein's *Tractatus Logico-Philosophicus*", *Mind* 32, 465-78.
- Suárez, Alfonso (2014) "The Metaphysical Status of the Objects of Wittgenstein's *Tractatus*", *Teorema* 33, 29-44.
- Welty, Ivan (2021) "Incompatibility, Inconsistency, and Logical Analysis in *Tractatus Logico-Philosophicus*", *Synthese* 199, 8171-86.
- Wittgenstein, Ludwig (1929) "Some Remarks on Logical Form", *Proceedings of the Aristotelian Society*, Supplementary Volume 9, 162-71.
- Wittgenstein, Ludwig (1975) *Philosophical Remarks*, edited by Rush Rhees and translated by Raymond Hargreaves and Roger White, Chicago: University of Chicago Press.
- Wittgenstein, Ludwig (1984) *Ludwig Wittgenstein und der Wiener Kreis*, edited by Brian McGuinness, Frankfurt: Suhrkamp.
- Wittgenstein, Ludwig (2005) *The Big Typescript*, edited and translated by C. Grant Luckhardt and Maximilian A. E. Aue, New York: Blackwell.
- Wittgenstein, Ludwig (2023) *Tractatus Logico-Philosophicus*, translated by Michael Beaney, New York: Oxford University Press.

Acknowledgement

Thanks to Timm Lampert, Gregory Landini, and Meredith Williams for comments on earlier drafts of this paper; to Dylan Fehr for proofreading; and to Riva Nayaju for help with the graphics.

Being Pleonastic

Paolo Bonardi (Vienna, Austria)

Abstract

In his book *The Things We Mean* (2003) and in other works, Stephen Schiffer has proposed an interesting and original theory of meaning. According to it, statements (or token declarative sentences) express *Schifferian* propositions, i.e. propositions which are (i) pleonastic, (ii) unstructured, and (iii) more or less fine-grained depending on the conversational contexts where the statements are made.

This paper will examine the feature (i) of Schifferian propositions. I will first introduce and attempt to elucidate the notions of pleonastic proposition and pleonastic entity more generally, by presenting five interpretations of them in terms of nonexistence, essence, ontological dependence, supervenience, and conceptual soundness. I will select the interpretation that seems to me most convincing and, in order to address certain challenges, I will propose refinements to those notions. Yet, additional problems will arise, to the extent that the usefulness of pleonastic propositions and pleonastic entities will be ultimately called into question.

In his book *The Things We Mean* (2003) and in other works, Stephen Schiffer has proposed an interesting and original theory of meaning. According to it, statements (or token declarative sentences, i.e. utterances and inscriptions of type declarative sentences) express *Schifferian* propositions, i.e. propositions which are (i) pleonastic, (ii) unstructured, and (iii) more or less fine-grained depending on the conversational contexts where the statements are made.

The present paper focuses on the feature (i) of Schifferian propositions. After introducing (§1), attempting to elucidate (§2), and even refine (§3) Schiffer's notions of pleonastic proposition and pleonastic entity more generally, it will be argued that these notions are problematic in various respects (§4, §5), to the extent that their usefulness will be ultimately called into question.

1. Pleonastic entities and pleonastic propositions

Schiffer (2003) characterizes pleonastic entities, viz. properties and propositions, as follows:

Pleonastic entities are entities whose existence is typically secured by something-from-nothing transformations We have a something-from-nothing transformation when from a statement involving no reference to

an *F* we can deduce a statement that does refer to an *F*. The property of being a dog is a pleonastic entity. From the statement

Lassie is a dog,

whose only singular term is “Lassie”, we can validly infer the pleonastic equivalent

Lassie has the property of being a dog,

which contains the new singular term “the property of being a dog”, whose referent is the property of being a dog. (2003: 61)

Incidentally, by “valid inference” Schiffer means *conceptually* (rather than logically) valid inference (2016: 390).

Propositions ... are also pleonastic entities. They have their something-from-nothing transformations, such as the one that takes us from

Lassie is a dog,

whose only singular term continues to be “Lassie”, to another of its pleonastic equivalents,

That Lassie is a dog is true

... which contains the singular term “that Lassie is a dog”, whose referent is the proposition that Lassie is a dog. (2003: 71)

In addition to properties and propositions, the list of pleonastic entities includes fictional characters (2003: 51), events (2003: 63) and arguably all abstract entities.

So, like Frege, Russell and their contemporary followers, Schiffer admits abstract entities, but he regards them as *pleonastic*. This is thought to be an advantage of Schiffer’s view over the Fregean/Russellian view: the former should be ontologically more parsimonious or less committed than the latter

[We should not] take the existence of [pleonastic] propositions very seriously. They exist, but only in a very *deflationary*, or minimalist ... way. (1990: 268 – boldface mine)

[Pleonastic entities] are hypostatizations of certain linguistic or conceptual practices. (2000: 7)

[A]s properties are shadows of predicates, so propositions are shadows of sentences. (2003: 71)

Yet, it is somehow unclear what it means for something *to exist in a deflationary or minimalist way*, or *to be a hypostatization or a shadow of something else*.

2. What exactly does “pleonastic” mean?

First hypothesis: *nonexistence*

It is rather common among analytic philosophers nowadays to conceive *existence* as a *first-order* property (Salmon 1987; Kaplan 1989: 541 and 1989a: 580n29; Crane 2013: 34; Priest 2016: 13, 59), especially after Salmon’s (1987) influential arguments in favor of it. Based on this conception, the totality of all objects can be divided into two (mutually exclusive and jointly exhaustive) categories: *existent* objects, which exemplify the (first-order) property of existence; and *nonexistent* objects, which we can quantify over and refer to but which lack such a property. The passages by Schiffer (1990, 2000, 2003) cited at the end of §1 (especially the underlined sentence) may lead to hypothesize that pleonastic “entities” are nonexistent objects.

I think this hypothesis should be rejected. For, first, there is no evidence that Schiffer subscribes to the conception of existence as a first-order property and to nonexistent objects. Second, even in recent works such as (2016), he unequivocally affirms that something-from-nothing inferences establish the *existence* of pleonastic entities.

Second hypothesis: *essence*

Another option that must be turned down is one hypothesizing that the *essence* of a pleonastic entity, e.g. a proposition, is contained within the *essence* of the premise of a something-from-nothing transformation: the former essence, unlike the latter, includes e.g. the role of *primary* truth bearer with truth conditions that are *essential* and *absolute*.

Third hypothesis: *ontological dependence*

A further hypothesis posits that pleonastic entities *ontologically depend* on the premises of their something-from-nothing transformations.

Ontological dependence: An object *a* ontologically depends on an object *b* if and only if *a* can exist only if *b* exists.

In the following passage, Schiffer would seem to endorse such a hypothesis:

Unlike electrons, trees, rocks and other things that enjoy the highest degree of **ontological** and conceptual **independence** from our linguistic and conceptual practices, pleonastic entities ... [are] determined by our hypostatizing linguistic practices. (2000: 9 – boldface mine)

But in (2003) Schiffer clarifies:

Pleonastic entities are entities whose existence is typically secured by something-from-nothing transformations – “secured” not necessarily in the sense that they are brought into existence (like fictional entities) ... (2003: 61)

Of course, if pleonastic entities are *not brought into existence* by something-from-nothing transformations, then the former do not ontologically depend on (the premises of) the latter. Schiffer’s rejection of the ontological-dependence hypothesis is even more explicit in the following passage from (2016):

I do say that propositions, properties et al. are not as ontologically and conceptually independent of us as rocks and electrons, [but] I nowhere say that pleonastic entities other than fictional characters are language-created ... (2016: 392)

Regarding properties specifically, Schiffer offers an argument against the hypothesis under consideration:

I don’t see how [the view that pleonastic properties are creations of our conceptual or linguistic practices] can be literally true, since properties exist in every possible world, and thus in possible worlds in which there are neither thinkers nor speakers. (2003: 66)

Fourth hypothesis: *supervenience*

Pleonastic entities are entities whose existence is typically secured by something-from-nothing transformations – “secured” ... in the sense that their existence **supervenes** on the premises of something-from-nothing transformations. (2003: 61 – boldface mine)

I’m not sure to understand the claim that (j) *the existence of pleonastic entities* supervene on the premises of something-from-nothing transformations, at least if the standard characterization of supervenience below is adopted. Conversely, I understand, based on that characterization, the claim that (jj) *pleonastic entities* supervene on the premises of something-from-nothing transformations: (jj) entails that if two pleonastic entities differ, then the premises of their something-from-nothing transformations must also differ.

Supervenience: The *Fs* supervene on the *Gs* if and only if there can be a difference in the *Fs* only if there is a difference in the *Gs*.

However, the supervenience hypothesis raises questions about the purported advantage (highlighted in §1) of Schiffer's ontology over the Fregean/Russellian one: it is true that abstract entities à la Frege/Russell enjoy complete independence of language and mind, whereas pleonastic abstract entities would supervene on linguistic and cognitive practices; yet, *the latter entities, exactly as the former, would exist and would be ontologically independent of those practices*; consequently, the advantage of Schiffer's ontology over the Fregean/Russellian one would be very limited, if not insignificant.

The supervenience hypothesis also faces another problem. Suppose that by observing two similar dogs without knowing their names and without possessing sufficient descriptive information to discriminate them, Mary utters (1) below, once referring to one dog and another time referring to the other dog. Additionally, consider that according to Perry (1993), the demonstrative “it” in (1) is not replaceable by any coreferring term without compromising its cognitive significance. All this poses a challenge to the view that the pleonastic propositions expressed by Mary's utterances of (1) *supervene* on premises of something-from-nothing transformations: in this case, it seems that *two different* propositions are introduced using twice *one and the same* sentence, (1), which may play the role of the premise of a *unique* something-from-nothing transformation.

(1) It's a beautiful dog.

Fifth hypothesis: *conceptual soundness*

In personal communication, Schiffer has insisted that the existence of pleonastic entities, e.g. the pleonastic proposition *that Lassie is a dog*, is secured by the conceptual soundness of the inference (Inf 1) below. An inference is *conceptually sound* if and only if all of its premise are true and there is no conceptually possible world where all its premises are true and its conclusion is false.

(Inf 1)

(P1) Lassie is a dog.

So, (C1) that Lassie is a dog is true.

Nevertheless, as in the case of the supervenience hypothesis, it is unclear what the ontological advantage of the pleonastic conception of abstract entities over Frege's/Russell's conception would be.

Moreover, conceptual soundness relies on the notion of *concept*, which, in the context of Schiffer's theory, is a delicate notion: concepts, i.e. ways of thinking of objects (2016: 411, 433), might be regarded as pleonastic entities; of course, defining a pleonastic entity by appealing to pleonastic entities themselves risks to create a vicious circle.

3. Attempts to refine Schiffer's characterization of pleonastic entity

Restoring the ontological-dependence hypothesis

In order for the ontology of pleonastic abstract entities to be more parsimonious or less committed than the Fregean/Russellian one, it seems to me indispensable that the *ontological-dependence hypothesis* is adopted, pace Schiffer. As regards at this point the supervenience and the conceptual-soundness hypotheses, they can either (a) be integrated in some way with the ontological-dependence hypothesis or (b) be dismissed altogether. Option (b) seems to me the preferable one, given the difficulties posed by those two hypotheses.

Nonexistent pleonastic objects

However, once the ontological-dependence hypothesis is restored, the problem mentioned in §2 of properties that appear to exist in worlds where there are no speakers/thinkers re-arises. A solution to it could be attained by invoking the notion of nonexistent object (§2): worlds devoid of speakers and thinkers might be supposed to contain pleonastic properties and pleonastic objects more generally *as permanently nonexistent properties/objects*.

Incidentally, a similar strategy could also be invoked to account for pleonastic propositions which were intuitively true at a time when no speakers/thinkers had yet appeared in our spatiotemporal universe (e.g. the proposition *that immediately after the Big Bang, matter was very hot and dense*): we might say that those propositions were already “around” at that time but only as *nonexistent objects capable of bearing truth*. They came into existence (i.e. they acquired the first-order property of existence) at a later time when speakers/thinkers came into existence, more precisely when the premises of appropriate something-from-nothing transformations (or at least their parts) came into existence. The thesis that nonexistent propositions can bear truth is also upheld by Salmon (1998: 286) and Soames (et al. 2014: 102-103).

Something-from-nothing transformations as inscriptions

We have established that pleonastic entities depend (at least ontologically) on the premises of something-from-nothing transformations. Now, something-from-nothing transformations are inferences (§1), i.e. sequences of *type* (declarative) sentences. Type expressions are typically regarded as *abstract* artifacts, as such falling into a category of things that Schiffer regards as pleonastic. But if something-from-nothing transformations themselves are pleonastic, then the characterization of pleonastic entity leads to a vicious circle.

We might of course concede that something-from-nothing transformations are sequences of *token* (declarative) sentences, thus concrete entities. If so, a single pleonastic entity would not be associated to a single something-from-nothing transformation as Schiffer usually claims, but to a *class of tokens of a something-from-nothing transformation type*.

Even with this concession, though, the circle persists: by definition, a token sentence is either an inscription or an utterance; utterances are events and, on

Schiffer's (2003: 63) view, events are pleonastic entities. So, in order to avoid that pleonastic entities end up depending on other pleonastic entities, we must exclude utterances from the aforementioned class of something-from-nothing transformation tokens.

Summing up, *a pleonastic entity would be an entity that depends, at least ontologically, on a class of inscriptions of the premise of a something-from-nothing transformation type*. In the remainder of the paper, I will show that even this refined characterization of pleonastic entity encounters obstacles.

4. A difficulty with language-independent propositions

The view that pleonastic propositions depend on linguistic and cognitive practices (§3) clashes with Schiffer's general definition of proposition:

[propositions are] abstract, **mind- and language-independent** entities that have truth conditions, and have their truth conditions both essentially and absolutely. (2003: 14 – boldface mine)

In itself, this is not a serious problem: advocates of pleonastic propositions could simply discard Schiffer's definition of proposition or classify pleonastic propositions as *sui generis* propositions.

A more severe concern instead arises from the quasi-linguistic nature of pleonastic propositions. Consider sentence (2) below, where the "that"-clause refers to the proposition *that the food is ready*. Assume that (2) is true. Incidentally, propositions cannot be seen; (2) should in fact be construed as something like (2*) below (Kaplan, personal communication). Now, if propositions are pleonastic, namely they are "shadows of sentences" (Schiffer 2003: 71), it is doubtful that (2)/(2*) is true, since Lassie is a *languageless* creature. An objection of this sort is raised against logo-centric propositions by Soames (et al. 2014: 174, 176).

(2) Lassie sees that the food is ready.

(2*) Lassie knows that the food is ready, by seeing the food.

5. Doubts about something-from-nothing transformations

Do something-from-nothing transformations suffice to render entities pleonastic?

It is far from obvious that

(T) Being introduced by a something-from-nothing transformation is a sufficient condition for an entity to be pleonastic.

In fact, consider the something-from-nothing transformation (Inf 2) below, proposed by Amie Thomasson (2016). According to Thomasson, it follows from (Inf 2) and thesis (T), *unacceptably*, that tables are pleonastic entities.

(Inf 2)

(P2) There are particles arranged tablewise.

So, (C2) there is a table.

Schiffer (2016: 395), of course, denies that tables are pleonastic entities. However, he (2016: 393-394) rebuts Thomasson's objection by arguing that "arranged tablewise" abbreviates a phrase like the following:

so arranged that some subset of [those particles] would constitute a **table** if there are any tables. (2016: 393 - boldface mine)

Since this phrase, and consequently the premise (P2), encompasses the word "table", (Inf 2) does not qualify as a genuine something-from-nothing transformation.

I think there is a way to reformulate Thomasson's objection so as to circumvent Schiffer's reply. Consider a particular table, *t*, situated (at a fixed time) in a room, *r*, where no other tables are present. Let's reformulate (Inf 2) as (Inf 2*) below. Schiffer correctly contends that the phrase abbreviated by "tablewise" includes the word "table". However, "table" is not a singular term referring to the table *t*: it is a common name referring to the kind *table* (Salmon 2012: 472). In fact, the abbreviated phrase contains *no* term referring to *t*.

(Inf 2*)

(P2*) There are particles arranged tablewise in *r*.

So, (C2*) there is *t* in *r*.

Perhaps, Schiffer could counter that even (Inf 2*) does not qualify as a genuine something-from-nothing transformation, because (P2*) contains a term, “table”, which refers to a *pleonastic* entity, viz. the *abstract* (Salmon 2012: 472) kind *table*.

Consider then (Inf 3) below. This inference should evade both objections raised above against (Inf 2) and (inf 2*): the phrase “collectively designated by an inscription ‘Vienna’” in (P3) does not encompass any term referring to Vienna or referring to any abstract (and thus pleonastic, on Schiffer’s view) entity. Therefore, (Inf 3) should qualify as a genuine something-from-nothing transformation, leading, on the basis of thesis (T) above, to the conclusion that Vienna is a pleonastic entity. Since this conclusion is *unacceptable*, (T) must be rejected.

(Inf 3)

(P3) There are particles collectively designated by an inscription “Vienna”.
So, (C3) there is Vienna.

In a passage of (2000), Schiffer seems to admit the falsity of (T), implicitly replacing it with something like (T*) below:

pleonastic entities ... have, as Mark Johnston (1988) would put it, “no hidden and substantial nature for a theory to uncover. All we know and all we need to know about [them] in general” is determined by our hypostatizing linguistic practices. (2000: 9)

(T*) Being introduced *and being exhaustively explained by* a something-from-nothing transformation is a sufficient condition for an entity to be pleonastic.

On the other hand, I suspect most metaphysicians would dispute the claim that entities such as propositions, attributes, fictional characters and events lack a “hidden and substantial nature for a theory to uncover” (Johnston 1988), which (claim) warrants the italicized insertion in (T*). Hence, (T*) is another dubious thesis.

Something-from-nothing transformations?

Reconsider Schiffer’s something-from-nothing transformation (Inf 1) at the end of §2. The pleonastic proposition *that Lassie is a dog* depends on the premise

(P1). Now, (P1) is more than a mere sequence of symbols: it is an English sentence, thus a sequence of symbols endowed with a meaning and thereby with a semantic content. So, the *pleonastic* proposition *that Lassie is a dog* depends on the sequence of symbols (P1) *plus its semantic content*. This content, i.e. what (P1) expresses, cannot be the *pleonastic* proposition *that Lassie is a dog* itself: otherwise, we would fall into a vicious circle. The content in question will be, by exclusion, a *non-pleonastic* proposition *that Lassie is a dog*. But, what is the point of introducing pleonastic propositions in ontology if their explanation ultimately depends on corresponding non-pleonastic propositions? This concern extends to all pleonastic entities: e.g. the pleonastic property of *being a dog* ends up depending on the non-pleonastic property of *being a dog*, thus rendering the former entity redundant and thereby dismissible.

References

- Crane, Tim (2013) *The Objects of Thought*, Oxford: Oxford University Press.
- Johnston, Mark (1988) "The End of the Theory of Meaning", *Mind and Language* 3(1), 28-42.
- Kaplan, David (1989) "Demonstratives", in: Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 481-563.
- Kaplan, David (1989a) "Afterthoughts", in: Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 565-614.
- Perry, John (1993) *The Problem of the Essential Indexical and Other Essays*, New York: Oxford University Press.
- Priest, Graham (2016) *Towards Non-Being*, 2nd edition, New York: Oxford University Press.
- Salmon, Nathan (1987) "Existence", *Philosophical Perspectives* 1, 49-108.
- Salmon, Nathan (1998) "Nonexistence", *Noûs* 32(3), 277-319.
- Salmon, Nathan (2012) "Generality", *Philosophical Studies* 161, 471-481.
- Schiffer, Stephen (1990) "The Mode-of-Presentation Problem", in: Anthony Anderson and Joseph Owens (eds.), *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, Stanford: CSLI Publications, 249-268.
- Schiffer, Stephen (2000) "Pleonastic Fregeanism and Empty Names", in: Anthony Everett and Thomas Hofweber (eds.), *Empty Names, Fiction and Puzzles of Non-Existence*, Stanford: CSLI Publications, 3-16.

Schiffer, Stephen (2003) *The Things We Mean*, New York: Oxford University Press.

Schiffer, Stephen (2016) "Replies to the Essays", in: Gary Ostertag (ed.), *Meanings and Other Things: Themes from the Work of Stephen Schiffer*, New York: Oxford University Press, 385-527.

Soames, Scott; King, Jeffrey; Speaks, Jeff (2014) *New Thinking about Propositions*, New York: Oxford University Press.

Thomasson, Amie (2016) "Easy Ontology and its Consequences", in: Gary Ostertag (ed.), *Meanings and Other Things: Themes from the Work of Stephen Schiffer*, New York: Oxford University Press, 34-53.

Acknowledgement

I am grateful to Naoya Fujikawa for comments on a previous draft of this paper and to Damian Aleksiev and Stephen Schiffer for discussion on topics related to the paper. My research is currently funded by the *Austrian Science Fund* (FWF) [10.55776/M3350].

Are Artificial Neurons Neurons?

Johannes Brinz (Osnabrück, Germany)

Abstract

The media often discuss artificial neural networks like ChatGPT or Amazon's Alexa, and policymakers grapple with regulating emerging technologies. However, the precise nature of "artificial neurons" remains ambiguous. Is this term to be understood merely metaphorically or does it refer to physical entities resembling biological neurons? While commonly understood as mathematical nodes in AI, the discussion extends deeper, particularly with the advent of neuromorphic engineering. This paper discusses whether artificial neurons are indeed neurons and what the potential implications are. Specifically, the paper addresses the question: *Are artificial neurons functional analogs of biological neurons?* I discuss three types of artificial neurons against the background of the two main theories of biological function, viz. Causal Role and Selected Effects Theory. My thesis contends that while artificial neurons lack the morphological structure of biological neurons, certain types, especially those in neural circuits, exhibit functional parity, and thus can be considered functional analogs.

1. Introduction

The news are filled with talks of artificial neural nets such as ChatGPT or Amazons' Alexa, and politics currently struggle finding the right measures to cope with those emerging technologies. But what do we actually mean, when we speak of *artificial neurons*? Do we refer to real neurons (or something similar) as in "artificial hearts" or are we speaking metaphorically? Most people familiar with AI might think the answer is obvious: When we say, for example, that ChatGPT has over 100 billion "neurons" we are talking of mathematical nodes, or learning parameters, nothing alike biological neurons. In that sense artificial neurons have nothing to do with real neurons, just as bell curves have nothing to do with real bells. But I think, the question runs deeper than that. First, because all artificial neurons are implemented in hardware, and therefore become physical entities that do something essentially neuronal: processing information. Second, computers no longer need to be totally unlike biological brains. The field of *neuromorphic engineering* works on building chips that consist of actual physical neurons and synapses that operate with mechanisms akin to those of biological brains. So the question I want to be asking is: *Are artificial neurons neurons?*

But why would that be an interesting question to ask? First of all, with AI increasingly penetrating our everyday world, we should be clear about what exactly we mean when we use central terms, one of them being "artificial

neuron”. Is there a reason for calling them that way, other than it being a useful metaphor? The term “neuron” in our everyday use of language refers to very different structures that should be distinguished carefully in order to better understand what we are talking about, and in order to see whether those different structures might have some relevant similarities. Second, it has been argued elsewhere (Brinz, forthcoming) that systems consisting of actual neurons are more likely to generate artificial consciousness. When we implement artificial neural networks on neuromorphic, i.e. brain-like, hardware, we go from simulating towards replicating the brain. This might have important moral, legal and political implications (Metzinger, 2021; Gibert and Martin, 2022; Ladak, 2023; Shevlin, 2021; Gordon and Pasvenskiene, 2021).

In the present paper I argue that *artificial neurons are not real neurons*, since they are morphologically distinct from biological neurons, but that *some of them, however, are neural analogs, i.e. they serve the same biological function as real neurons*. I proceed as follows: First, I distinguish between three types of artificial neurons: (1) Implementations of mathematical neurons on standard digital hardware, (2) cores on neuromorphic chips, and (3) artificial neural circuits. Then, I discuss the two main positions in the philosophy of biological functions: Causal Role and Selected Effects theory. I conclude that neurons of category (3) come closest to being functional analogues of biological neurons.

2. Three categories of artificial neurons

As already mentioned, it makes sense to distinguish three categories of artificial neurons ranked by their degree of biofidelity. (1) First, let us consider implementations of artificial neural networks (ANNs) in standard digital hardware. ANNs are abstract mathematical models that describe high dimensional computations with graphs of interconnected nodes, so called “neurons”. Effectively, they utilize matrices and non-linear functions to process and transform input data, enabling them to learn complex patterns and relationships through training. This allows them to perform various tasks such as classification, regression, and pattern recognition in machine learning applications. In this context “artificial neuron” refers to parts of graph theoretical descriptions of complex computations, thus to *abstract* entities. However, in actual AI applications ANNs are always implemented in some

hardware. The most common hardware architecture is the von Neumann architecture that consists of a CPU, a memory device, input/output devices, etc. Since the vast majority of computers today operate according to this hardware design, I refer to it as standard digital hardware (SDH). On a very simplified view, artificial neurons in SDH are certain numbers (8, 16, 32, etc.) of capacitors that change their state (charged, discharged) according to the program specified by the ANN, and are arranged in a grid-like manner. (2) Second, of particular interest to the question at hand is a different kind of hardware architecture. Neuromorphic hardware comprises a range of different computational approaches, all of which have in common that they consists of “neurons”, i.e. circuits that emulate the electrical behavior of biological neurons up to a certain degree of accuracy and “synapses” that connect those neurons with a variable conductivity. Examples include Intel’s Loihi (Davies et al., 2018), IBM’s TrueNorth (Merolla et al., 2014), and the very large scale system SpiNNaker2 (Mayr & Furber 2019). Those systems are developed to simulate ANNs with higher speed and energy efficiency. (3) The third category consists of small scale research chips that use the same organizing principles as the nervous systems of biological brains. Examples range from sensory systems (Wen and Boahen 2009), to networks with biologically plausible neural dynamics (Benjamin et.al. 2014)), to spike based learning circuits (Qiao et.al. 2017). For further detail see (Indivieri 2021). Some systems can even be used as neuroprotheses (Abu-Hassan et.al. 2019).

3. The biological function of a neuron

In neuroscience the textbook definition of neurons is something like: “cells specialized for the generation, conduction, and transmission of electrical signals [...]” (Purves et.al. 2018), or: “the cell type that conveys information.” (Delcomyn 1998) Thus, neurons are partly defined morphologically as a certain type of biological cell. This precludes all three types of artificial neurons from being real neurons. Therefore, the answer to the question: *Are artificial neurons neurons?* strictly speaking must be: *No*. However, in biology the concept of (functional) analogy exists, i.e. the principle that two morphologically different structures can have the same biological function (Abouheif et.al. 1997, Amundsen & Lauder 1994). A well known example are isofunctional enzymes, i.e. structurally different molecules that

catalyze the same chemical reaction in the human metabolism. (Piergiore et.al. 2017) So the more interesting question is: *Are artificial neurons functional analogues of biological neurons?*

To answer that question, we first need to understand what the function of a neuron is. In philosophy of biology there are two main approaches to biological functions: Causal Role, and Selected Effects Theory. I discuss them consecutively.

The **Causal Role Theory** of biological functions is usually attributed to Cummins (1975 and 1983, ch. 2), where he tries to define the function of a certain trait. If we translate his definition to the case of neurons and the brain we get the following statement:

[A neuron] functions as a φ in [the brain] (or: the function of [a neuron] in [the brain] is to φ) relative to an analytical account A of [the brain]'s capacity to ψ just in case [the neuron] is capable of φ -ing in the brain and A appropriately and adequately accounts for [the brain]'s capacity to ψ by, in part, appealing to the capacity of [the neuron] to φ in [the brain]. (Cummins 1975: 762)

So the question now is: What is φ ? What is the function of the neuron? According to Cummins, in order to answer that question, we first need to know what ψ is. Which capacity of the brain is it that neurons contribute to? I think three capacities should be considered: (ψ_1) information processing, (ψ_2) information integration, (ψ_3) electrical signaling. Neurons might have different functions that contribute to those different capacities of the brain. Cummins is also not particularly clear on what an analytical account is supposed to be. Following that critique Carl F. Craver (2000) proposed that systems in Cummins account should be understood as mechanisms and that “[a]n analytic account for a mechanism is not just a list of entities and activities; [...] it involves, in addition [...], a description of how they are organized together actively, spatially, and temporally [...].” (Craver 2000: 61) If we want to understand what the mechanistic role, i.e. the function, of a neuron is we need to know how exactly it fits into the causal mechanisms of the brain. “It is by

detailing how an item fits into the spatial, temporal, and active organization of a mechanism (showing exactly how it contributes to S's ψ -ing) that one specifies its mechanistic role." (Craver 2000: 62)

On this understanding, the function of a trait is its disposition to bring about a certain effect to a given cause in a specific context (Walsh & Ariew 1996). Therefore, any two objects with the same function respond to the same cause with the same effect as described by a given account. Any functional analog must have the same disposition to respond to a given input stimulus (cause) in a certain way (effect). This fact opens up the possibility of what I call the neuronal replacement test of (biological) function:

Given a capacity ψ , a description A , and a part π of a mechanism, an item π' is an functional analog of π if and only if the system where π is replaced with π' does still bring about ψ as a result of the mechanism described by A .

Now we can apply the replacement test to neurons and brains. Let us begin with the capacity to produce electrical signals (ψ_3) and a standard neurobiological explanation of it. A given entity then is a functional analog of a certain neuron, if the latter can be replaced by the former without affecting the neurobiological mechanisms that produce the electrical signals of the brain. So for ψ_1 the application of the replacement test is rather straight forward: Replace one biological neuron with an artificial one and see if the brains electrical mechanisms are still intact. If they are and the brain sends the correct impulses to muscles and glands, the artificial neuron is a functional analog with respect to electrical signaling.

One main criticism of the Causal Role Theory, issued by the **Selected Effects Theory** (Wright 1973, Milikan 1987), is that it cannot distinguish between actual functions and mere side-effects. Besides pumping blood, one of the effects of hearts is making throbbing noises that leave marks on an cardiogram. According to the Causal Role Theory, heart, body, and cardiogram form a mechanism, and the causal role of the heart in that mechanism is to make throbbing noises which the cardiogram picks up. But it seems wrong to say that the function of the heart is to make throbbing noises. So what is special about pumping blood? The answer of the Selected Effects Theory is that pumping blood is the one effect of the heart that it was selected for in the

evolutionary history. Healthy animals survived because their hearts were pumping sufficient amounts of blood, not because they were making certain noises. Making throbbing noises is a mere side-effect of the hearts actual function, viz. pumping blood.

From this perspective, one might wonder whether something similar is true for the functions of the brain. It is alluring to say that processing (ψ_1) and integrating information (ψ_2) are the relevant functions of the brain while electrical signaling is only a mere side-effect. It is the fact that the information about dangerous animals and nourishing food sources were processed and integrated correctly that gave animals with healthy brains an evolutionary advantage. The fact that this process is implemented in an electrical neuronal mechanism seems to be merely accidental. I believe this objection holds. Neural encoding is only one out of many ways to process and integrate the relevant information about danger and food. However, so far we have not been discussing a fourth important capacity of the brain: (ψ_3) Consciousness. Consciousness is generally considered not only a side-effect but itself a function of the brain. Arguments have been put forward that conscious thought production yields an evolutionary advantage (Eccles 1992). While computationalists claim that processing information in the right way is sufficient for a system to have mental states (Rescorla 2015), and e.g. Integrated Information Theory conjectures that integrating information in similar ways means having similar conscious experiences (Tononi 2004), others believe that neither processing, nor integrating information is sufficient for consciousness. They believe that the specific neural mechanisms give rise to biological (Smart 2022) and artificial consciousness (Gamez 2020). Under the assumption that consciousness is grounded in the processes described by current neuroscience, the mechanisms that account for the electrical signaling are the same that generate conscious experience of biological brains.

4. Results

If it is correct that the brains functions are either (ψ_1) information processing, (ψ_2) information integration, (ψ_3) conscious thought production, or a combination of those three, then we can now derive the biological functions of

neurons. The function of neurons is either (φ_1) being the appropriate operational component in a certain information processing system, (φ_2) integrating information the right way, (φ_3) responding to incoming spike-trains from the post-synaptic side with the appropriate pattern of action potentials on the pre-synaptic side, or a combination of those three. I now consider the three types of artificial neurons to see if they can serve as functional analogs for the different neuronal functions (φ_1 to φ_3). Let's begin with SDH. If we have an appropriately programmed ANN SDH neurons can be used as operational components in a system that processes information in the same way as a biological brain does (φ_1). However, they are neither capable of integrating information appropriately (φ_2) (Koch 2019, Tononi and Koch 2015), nor do they produce any spiking potentials (φ_3). Neuromorphic chips also do not have the same spiking behavior as biological neurons (φ_3), however, they are capable of processing (φ_1) and also integrating information in a brain-like manner (φ_2) (Koch 2019, Tononi and Koch 2015). Artificial neural circuits are the most similar to biological neurons. They are in principle capable of processing (φ_1) and integrating information (φ_2) in the right way and they can be build to have the same spiking behavior as biological neurons (φ_3). Some artificial neurons are specifically designed to pass the neural replacement test, i.e. they are build as neuro-protheses (Abu-Hassan et.al. 2019). These results are summed up in table 1.

Table 1: Artificial neurons ordered by the question of whether they serve the corresponding function.

	Information processing (φ_1)	Information integration (φ_2)	Spiking behavior (φ_3)
SDH	✓	#	#
Neuromorphic cores	✓	✓	#
Artificial neural circuits	✓	✓	✓

5. Conclusion

In the present paper I argued that artificial neurons cannot be regarded real neurons since they lack the right morphological structure. However, some of them can possibly be considered functional analogs, i.e. morphologically different entities that serve the same function. I differentiated between three types of artificial neurons: Neurons in SDH, neuromorphic cores, and artificial neural circuits. Considering the Causal Role and the Selected Effects Theory of biological function, I worked out what I believe to be the most relevant functions of neurons: Information processing, information integration, and electronic signaling. I then tried to show that SDH neurons only can serve the function of processing information, while neuromorphic cores additionally are capable of integrating information in the right way. Artificial neural circuits can serve all three functions of biological neurons, and thus are good candidates for being considered functional analogs.

References

- Abouheif, E., Akam, M., Dickinson, W. J., Holland, P. W. H., Meyer, A., Patel, N. H., Raff, R. A., Roth, L. V., & Wray, G. A. (1997). Homology and developmental genes. Abu-Hassan, K., Taylor, J. D., Morris, P. G., Donati, E., Bortolotto, Z. A., Indiveri, G., ... & Nogaret, A. (2019) "Optimal solid state neurons", *Nature Communications*, 10(1), 5309.
- Amundson, R., & Lauder, G. V. (1994) "Function without purpose: The uses of causal role function in evolutionary biology", *Biology and Philosophy*, 9, 443-469.
- Benjamin, B. V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A. R., Bussat, J. M., ... & Hasler, P. E. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, 102(5), 699-716.
- Brinz, J. (forthcoming). Neuromorphic AI – From Simulating Towards Replicating the Brain.
- Carrara, M., & Vermaas, P. E. (2009) "The fine-grained metaphysics of artifactual and biological functional kinds", *Synthese*, 169, 125-143.
- Craver, C. F. (2001) "Role functions, mechanisms, and hierarchy", *Philosophy of Science*, 68(1), 53-74.
- Cummins, R. E. (1975) "Functional analysis", *Journal of Philosophy*, 72, 741-764.
- Cummins, R. (1983) *The nature of psychological explanation*. MIT Press.

- Davies, M., Srinivasa, N., Lin, T., China, G., Cao, Y., Choday, S. H., ... & Tse, J. (2018) "Loihi: A neuromorphic manycore processor with on-chip learning", *IEEE Micro*, 38(1), 82–99.
- Eccles, J. (1992) "Evolution of consciousness", *Proceedings of the National Academy of Sciences*, 89(16), 7320-7324.
- Gamez, D. (2020) "The relationships between intelligence and consciousness in natural and artificial systems", *Journal of Artificial Intelligence and Consciousness*, 7(01), 51-62.
- Godfrey-Smith, P. (1993) "Functions: Consensus without unity", *Pacific Philosophical Quarterly*, 74, 196-208.
- Ielmini, D., Wang, Z., & Liu, Y. (2021) "Brain-inspired computing via memory device physics". *APL Materials*.
- Indiveri, G. (2021) "Introducing 'neuromorphic computing and engineering'", *Neuromorphic Computing and Engineering*, 1(1), 010401.
- Koch, C. (2019) "Proust among the machines", *Scientific American*, 321(6), 46-49.
- Mahowald, M., & Douglas, R. (1991) "A silicon neuron", *Nature*, 354(6354), 515-518.
- Mayr, C., Hoepfner, S., & Furber, S. (2019). "Spinnaker 2: A 10 million core processor system for brain simulation and machine learning". arXiv preprint arXiv:1911.02385.
- Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., ... & Modha, D. S. (2014) "A million spiking-neuron integrated circuit with a scalable communication network and interface", *Science*, 345(6197), 668–673.
- Millikan, R. G. (1987) *Language, thought, and other biological categories: New foundations for realism*. MIT press.
- Piergiorgio, R. M., de Miranda, A. B., Guimaraes, A. C., & Catanho, M. (2017) "Functional analogy in human metabolism: enzymes with different biological roles or functional redundancy?" *Genome Biology and Evolution*, 9(6), 1624-1636.
- Qiao, N., Bartolozzi, C., & Indiveri, G. (2017) "An ultralow leakage synaptic scaling homeostatic plasticity circuit with configurable time scales up to 100 ks", *IEEE Transactions on Biomedical Circuits and Systems*, 11(6), 1271-1284.
- Rescorla, M. (2015) "The computational theory of mind", in: E. N. Zalta & U. Nodelman (Eds.) *The Stanford Encyclopedia of Philosophy*, (Fall 2020)
- Smart, J. J. C. (2022) "The Mind/Brain Identity Theory", in: E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022).
- Tononi, G. (2004) "An information integration theory of consciousness" *BMC neuroscience*, 5, 1-22.

Tononi, G., & Koch, C. (2015) "Consciousness: here, there and everywhere?", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), 20140167.

Walsh, D. M., & Ariew, A. (1996) "A taxonomy of functions", *Canadian Journal of Philosophy*, 26(4), 493-514.

Wen, B., & Boahen, K. (2009) "A silicon cochlea with active coupling", *IEEE Transactions on Biomedical Circuits and Systems*, 3, 444–455.

Wright, L. (1973) "Functions", *Philosophical Review*, 82, 139-168.

Wright, L. (1976) *Teleological Explanations*. Berkeley: University of California Press.

Acknowledgement

I would like to acknowledge the support of the broader academic community and the resources provided by the University of Osnabrück that facilitated the development of this paper. I extend my gratitude to Leonie Buning and Louis Loock for their meticulous proofreading of this paper. Their keen attention to detail and valuable feedback have significantly improved the clarity and quality of the final manuscript.

Wittgenstein's *Tractatus*: A Transcendental Account

Cecilia Brotzu (Pavia, Italy)

Abstract

My essay aims to show that in Wittgenstein's writings, as early as in the *Tractatus logico-philosophicus*, there is an underlying idea of philosophy that can be traced back to a transcendental approach, understood in a Kantian way. Wittgenstein's intention consists in not only to deconstruct grammatical deceptions, but also defining its limits and field of legitimacy. In my article, the issues of the sentence as a picture and the representation are introduced, together with the consequent anti-realistic semantics; furthermore, the transcendental aspects of the *Tractatus* are explored; in light of Stenius' theoretical proposal, that takes shape as a discussion of the fundamental assumptions of the *Critique of the pure Reason*, on the basis of which the central theses of the *Tractatus* are examined, an analysis of Wittgenstein's work will be suggested, with particular attention to the theme of transcendental deduction. In closing, the problem of the ego is investigated, emphasizing that the "Mystical" does not constitute the exoteric part of Wittgenstein's masterpiece, but it is attributable to a transcendental perspective.

The purpose of my essay is to explore transcendental features in Ludwig Wittgenstein's *Tractatus logico-philosophicus*, by highlighting that the meaning of Wittgenstein's critique of language is ascribable to a Kantian idea of philosophy. Wittgenstein's intention consists in not only to deconstruct grammatical deceptions, but also defining its limits and field of legitimacy. The *Tractatus* does not directly focus on the language, rather on its conditions of possibility (cf. Borutti 2010: 36). In his preface to the work, Wittgenstein himself points out that «the book will, therefore, draw a limit to thinking, or rather, not to thinking, but to the expression of thoughts; for, in order to draw a limit to thinking we should have to be able to think both sides of this limit (we should therefore have to be able to think what cannot be thought)» (TLP: 27).

To ensure that the theoretical groundwork being analyzed is properly framed, the notion of image (*Bild*), central in the *Tractatus logico-philosophicus*, will be now examined; the conception outlined by Wittgenstein is known as representational theory. The defining factor of an image is the coordination of its elements with objects and their proper placement (cf. TLP 2.131), that «are combined with one other in a definite way» (TLP 2.14) and the image shares with its depiction what Wittgenstein calls "form of representation" (*Form der Abbildung*), which «is the possibility that the things are combined with one

another as are the elements of the picture» (TLP 2.151). In section 2.221, Wittgenstein calls “sense” (*Sinn*) what an image represents (*stellt dar*). Every image must have something in common with reality, and this is what Wittgenstein refers to as logical form (cf. Perissinotto 2018: 79). Section 2.18 states that «what every picture, or whatever form, must have in common with reality in order to be able to represent it at all, rightly or falsely, is the logical form, that is, the form of reality». Every image has its logical form; otherwise, its representative capacity would fail. Proposition 3.13 asserts: «to the proposition belongs everything which belongs to the projection; but not what is projected. Therefore, the possibility of what is projected but not this itself». M. Black underlines that the expression “everything that belongs to the proposition” must be read as everything that is *internal* to the depicting relationship, that means the logical form that the proposition has in common with what it represents (cf. Black 1964: 104). In section 4.12 it is stated that «propositions can represent the whole reality, but they cannot represent what they must have in common with reality in order to be able to represent it, the logical form. To be able to represent the logical form, we should have to be able to put ourselves with the proposition outside logic, that is outside the world». As spatial image cannot depict its own form of representation, similarly a proposition cannot represent what it must have in common with reality. «The picture represents [*stellt dar*] a possible state of affairs in logical space» (TLP 2.202). The depicting relationship is termed by the German verb *darstellen*, which means to represent, in the sense of presenting intuitively, also in a graphical form. While in the *Tractatus* the term *vorstellen*, to represent in the sense of an idea, concept or mental reproduction of sensory data, rarely occurs, the prevalence of the verb *darstellen* suggests that in Wittgenstein’s idea of linguistic representation, and, therefore, his semantic framework concerning the language-world relationship, the linguistic element of presentation through a form would prevail over its conceptual aspect (cf. Borutti 2010: 77). An image depicts a state of affairs (*Sachverhalt*); it does not represent a fact that actually occurs, but its possibility, since a proposition can be false and thus present a situation that does not exist in the world. The conception of language as image of the world implies a reinvention of the theme of image, of which Wittgenstein provides a non-mimetic idea (ivi: 75).

The *Tractatus* offers a mere logical characterization of the notion of thought, not referring to the psychological sphere. Wittgenstein does not reduce the field of thinkable to what can be represented through a mental image, but to what is capable of representation. By virtue of its depicting capability, an image is something which truth and falseness can be predicated of. In the proposition 4.001 it is declared that «the totality of propositions is the language». Wittgenstein does not aim to counterpose our common language with an ideal or logically perfect language. The essence of a proposition is not something to which our propositions should aim and to which they might, therefore, fail to correspond (cf. Perissinotto 2018: 72). Furthermore, a proposition can be understood regardless of whether it is true (cf. TLP 4.024).

The conception of image previously explained has shown how the sense of a proposition is internal, in a perspective of autonomy of language, or immanence of language, which implies that language says what it says on the basis of its own resources and not on the ground of conformity or correspondence to an external structure. Although the idea of a proposition as an image of a state of affairs could allude to a realistic semantics, the semantics of the *Tractatus* does not reduce sense to reference to objects in the external world, but is based on the autonomy of language, on its internal link with the reality represented (on this topic, see Appelqvist 2023: 2). Logic marks the limit of language; however, to be transcendental is not the formal logic, rather the logical analysis of language, which leads to recognition of the logic that underlies to the whole reality. In this particular sense, formal logic and transcendental logic end up overlapping. Formal logic thus also becomes transcendental, since, given that sense is conveyed in ordinary language, it is possible to trace condition of possibilities in the assumption that there is an isomorphism between language, thought and reality. As a result, the logical analysis can fulfill its transcendental function of detecting the conditions of sense, precisely because it presupposes something without which it could not operate at all (cf. Gargani 2003: 30).

In this regard I will follow the interpretation provided by E. Stenius, that takes shape as a discussion of the fundamental assumptions of the *Critique of the pure Reason*, on the basis of which the central theses of the *Tractatus* are examined. It starts from the consideration that Kant's question about how synthetic a priori judgments are possible is capable of establishing a

dichotomy between what belongs to the domain of theoretical reason and what does not. The mistake in Leibniz-Wolffian metaphysics, dogmatic according to Kant, consisted of illegitimately applying the forms of theoretical reason to questions lying outside its domain. This question serves as the basis for setting the *limits* of theoretical reason. The analysis provided by Stenius aims to show how some fundamental assumptions in Kant's thought can still be valid within a Wittgensteinian perspective. The Kantian thesis according to which a world is a world of possible experience only if it is possible to theoretical reason, that is, whether it is imaginable or intelligible, can also be accepted in Wittgenstein's philosophy, provided that imaginable and intelligible correspond to thinking and that thought is the logical image of reality, hence what is thinkable is what can be presented by a logical image, that can be described by a depicting language. Thus, being possible to the theoretical reason corresponds, in a Wittgensteinian perspective, to the possibility in terms of what is describable in a meaningful language.

This is the essential modification of the Kantian view which gives rise to all differences between Wittgenstein and Kant. The task of (theoretical) philosophy is for Wittgenstein as for Kant to indicate the limits of theoretical discourse. But since what belongs to theoretical discourse is what can be "said" at all in language, the investigation of this limit is the investigation of the "logic" of language, which shows the "logic of the world". (Stenius 1960: 218).

Given that the task of theoretical philosophy consists in a transcendental deduction that deals with the limits of theoretical discourse rather than speculations that transcend this limit and cannot be known theoretically, a milestone in Kantian thought, can also be accepted in a Wittgensteinian theoretical horizon, as long as we keep in mind that what belongs to theoretical discourse is what can be said in language (*ibidem*). Section 6.13 states: «logic is not a theory but a reflection of the world. Logic is transcendental»; according to Stenius, it can be interpreted as follows: «what Kant's transcendental deductions are intended to perform: this is performed by the logical analysis of language» (ivi: 220). The Kantian thesis according to which our experience has a form, grounded in theoretical reason, and a content, based on our senses, remains true in Wittgenstein's system, even though also the external structure of reality must be included in this content.

According to Stenius, the logical analysis of language as conceived by Wittgenstein is a kind of transcendental deduction in the Kantian meaning, whose purpose is to indicate the a priori form of experience, which is *shown* and cannot be *said*. Kantian philosophy can be defined as subjective, but only in a transcendental sense and non-empirical. Kantianism has been termed “Critical Idealism” or “Transcendental idealism”; similarly, «Wittgenstein’s philosophical system can be called “Critical Lingualism” or “Transcendental Lingualism” or even “Lingualistic Idealism”» (ivi: 220).

For Wittgenstein as well the form of experience is subjective in a foundational sense, in which the transcendental ego is to be distinguished from the empirical one. In this regard section 5.641 is significant: «there is therefore really a sense in which in philosophy we can talk of a non-psychological I. The I occurs in philosophy through the fact that the “world is my world”. The philosophical I is not the man, not the human body or the human soul of which psychology treats, but the metaphysical subject, the limit, not a part of the world». Therefore

The limit to thinking drawn by Wittgenstein’s transcendental deductions can, as we have seen, properly be drawn only in language. We cannot think what is unthinkable, but we can form linguistic expressions which do not express thoughts, since they are simply nonsensical. Thus the limit between what in Kantian terms belongs, and belongs not to theoretical reason is shown by the logical distinction between sense and nonsense. It follows that problems which according to Kant are unsolvable by theoretical reason cannot even be raised. (ivi: 222).

In line with Stenius’ work, K. O. Apel considers that, while Kant reframed the traditional problems of metaphysics in terms of conditions of experience, Wittgenstein transposed a “criticism of pure reason” into “criticism of pure language”. H. J. Glock shares a similar view; according to him, the most wide contact point between Wittgenstein and Kant must be recognized in their general understanding of philosophy, that both of them interpret as a work of delimitation of the field of scientific knowledge, and at the same time as an activity rather than a doctrine (cf. Bastianelli 2008: 73). Under the heading “philosophy” of his *A Wittgenstein Dictionary*, Glock specifies that the early Wittgenstein remains within the tradition of Kant’s critical philosophy and

places at the root of his vision the Kantian conception according to which philosophy reflects on the nature and precondition of representation. The “linguistic turn” undertaken by Wittgenstein, although in agreement with this Kantian idea, differs insofar as these conditions solely lie in a system of rules that regulates the logic of language. Wittgenstein draws a boundary Kantian-inspired between science, which pictures or represents the world, and philosophy, which reflects on the nature and conditions of possibility of this representation, as expressed in 4.11. M. Black points out that, in Wittgenstein’s work, the term transcendental must be understood as “beyond experience” and “a priori”; reporting an expression from Elizabeth Anscombe, he continues that it does not mean that logical propositions assert transcendental truths, but they show something which pervades the whole sphere of sayable, as all propositions do (cf. Black 1964: 329). On this topic also seems to be significant the section 5.552, according to which «the “experience” which we need to understand logic is not that such and such is the case, but that something *is*; but that is *no* experience. Logic precedes every experience, that something is *so*. It is before the How, not before the What».

Transcendental propositions, which, according to Kantian system cannot be *known* as true by theoretical reason but only *postulated* by practical reason, are not detectable in Wittgenstein’s perspective, as what they are trying to say cannot be said. However, the ineffable plays a role of primary importance; Wittgenstein says that «this shows itself; it is the mystical» (TLP 6.522). In Wittgenstein, unlike the positivistic nonsense, the unsayable is not a set of pseudo-propositions, but it bears a transcendental function: it is internally and necessarily linked with what can be said. Saying cannot be comprehensible if not opposed to the unsayable. In other terms, if it does not delimit it (cf. Borutti 2010: 168). Furthermore, Stenius notes that, unlike in a logical empiristic perspective, where the nonsense has a mere negative meaning, for Wittgenstein this notion reveals a positive value. However, the Mystical Wittgenstein talks about should not be understood as the esoteric aspect of his thought, as it is sometimes presented; Wittgenstein in fact had no inclination toward mysticism (*ibidem*). It rather constitutes the nucleus of a classical way of philosophizing, sharing traits with the transcendental perspective.

In order to secure the objectivity of the experience, Kant offers the solution of a universal subjectivity, which plays the role of a unifying center of the

multiplicity of experience. From Kant onwards, the subject has been configured not as a mere mirror of an independent reality, but as a true core of knowledge and guarantee of the stability of an objective and intersubjective real world. In other words, the idea of knowledge consisting of a mere reflection of an already given reality has been abandoned. In the *Tractatus* the world is represented through meaningful propositions within the limits of logical space. Therefore, logic is the all-pervasive medium in which the human experience of reality takes place and the transcendental research becomes an investigation into the condition of meaning of language. The result is that representational knowledge is made possible by the original isomorphism between reality and thought, manifested in the logical form of propositions (cf. Bastianelli 2008: 173).

The logic at the core of Wittgenstein's theory of representation is, at first glance, a logic without subject. The risk of falling back into a solipsistic perspective is made clear by the well-known section 5.6: «the limits (*Grenzen*) of my language mean the limits of my world». However, the anti-realistic semantics underlying to the theory of language as representation does not imply that world and language coincide. Regarding the solipsistic risk inherent in the *Tractatus*, see the section 5.633: «*Where in the world is a metaphysical subject to be noted? You say that in this case is altogether like that of the eye and field of sight. But you do not really see the eye. And from nothing in the field of sight can it be concluded that it is seen from an eye*». And also, in 5.634: «this [the fact that the visual field cannot be traced back to the eye] is connected with the fact that no part of our experience is also a priori. Everything we see could also be otherwise. Everything we can describe at all could also be otherwise. There is no order of things a priori». An eye, while watching, cannot see itself. It is not from elements contained in the visual field that we understand we have eyes. The argument about the visual field shows that we are not talking about an empiric subject which experiences the world and represents it; rather, we are talking about language as an opening and at the same time a limit to the world. Black points out the fact that each empirical subject experiencing his own world is not a contingent fact. Therefore, if the metaphysical subject were identifiable within experience, there would be nothing a priori to discover as part of experience.

The final section of the *Tractatus*, probably the most renowned of the work, says that «whereof one cannot speak, thereof one must be silent». (TLP, 7). Consistently with the assumptions of the work already discussed, this duty is not to be understood in a deontic sense, rather as the impossibility in principle of saying something that cannot be said, showing from an inner perspective the limits of language and the world.

References

- Appelqvist, Hanne (2023) *Wittgenstein and Aesthetics*, Cambridge: Cambridge University Press.
- Bastianelli, Marco (2008) *Oltre i limiti del linguaggio. Il kantismo nel Tractatus di Wittgenstein*, Milano: Mimesis.
- Black, Max (1964) *A Companion to Wittgenstein's Tractatus*, Ithaca: Cornell University Press.
- Borutti, Silvana (2010) *Leggere il Tractatus logico-philosophicus di Wittgenstein*, Como-Pavia: Ibis.
- Frascolla, Pasquale (2006) *Understanding Wittgenstein's Tractatus*, London: Routledge.
- Gargani, Aldo Giorgio (2003) *Wittgenstein. Dalla verità al senso della verità*, Pisa: Edizioni Plus – Università di Pisa.
- Glock, Hans-Johann (1996) *A Wittgenstein Dictionary* (1996), Oxford: Blackwell Publishers.
- Perissinotto, Luigi (2018) *Introduzione a Wittgenstein*, Bologna: Il Mulino.
- Stenius, Erik (1960) *Wittgenstein's Tractatus. A critical Exposition of the main lines of Thought*, Oxford: Blackwell Publishers.
- Wittgenstein, Ludwig (1922) *Tractatus logico-philosophicus*, London-Bradford: Lund Humphries.

Acknowledgement

I would like to thank Professor Serena Feloj (University of Pavia) for encouraging me to write this paper and Professor Silvana Borutti (University of Pavia), for always being a fundamental academic reference for Wittgenstein studies; special thanks also go to my dear friend Claudio, whose philosophical exchanges are always fruitful, and to Kerstin, for her patience and assistance with German language.

An Argument for Modal Realism

David Builes (New Jersey, USA)

Abstract

Extended Abstract. All facts are supposed to supervene on the fundamental facts. However, negative facts pose a problem for this supervenience thesis. If fundamental facts only include "positive" facts, concerning (i) what fundamentally exists and (ii) what fundamental properties and relations are had by what exists, then such facts don't seem capable of necessitating negative facts, such as "there are no ghosts". After all, it is consistent with the fundamental positive facts that there are also additional ghosts. The fundamental positive facts only say what *is*, they don't say what is *not*.

Alternatively, if the fundamental facts are all "atomic" (e.g. facts of the form [Fa] or [Rab], which do not utilize any logical ideology), then such facts also have don't seem to be able to necessitate negative facts like "there are no ghosts".

In light of this problem, most philosophers reject the view that all fundamental facts are positive and/or atomic. However, the goal of this talk will be defend both that (i) all fundamental facts are positive/atomic and (ii) all facts supervene on the fundamental facts.

I begin by first motivating the view that all fundamental facts are positive and atomic. Such a view can be motivated on the basis of parsimony concerns (admitting extra fundamental logical ideology is less parsimonious), anti-arbitrariness and anti-redundancy concerns (the question of which logical connectives are fundamental threatens to be either arbitrary or redundant), and on the basis of modal recombination arguments (fundamental positive facts and negative facts cannot be freely modally recombined). Moreover, I argue that standard accounts of "totality" facts do not fulfill the work that they are supposed to do.

I next turn to criticize some recent suggestions in the literature for how fundamental positive/atomic facts might serve as a supervenience base. In particular, I argue that a commitment to Monism (the view that there can only be one fundamental entity) does not solve the problem, and I argue that a commitment to Necessitism (the view that necessarily, everything is necessarily something) does not solve the problem. One common problem with both of these views is that, while they do manage to account for why there cannot be any extra fundamental entities, neither view is able to account for why there cannot be extra fundamental properties/relations had by the fundamental entities.

Next, I argue that a view similar to David Lewis' Modal Realism is able to resolve these problems. In particular, the following principle is needed:

Island Universe Plenitude: If it is possible for an island universe to have a maximally specific qualitative character Q, then there is such an island universe.

Crucially, unlike David Lewis' view, the view that I defend is *not* committed to the reductionist claim that modal facts hold in virtue of non-modal facts. It is also not committed to a Humean recombination principle across possible worlds.

Given **Island Universe Plenitude**, I argue that the fundamental facts can both be positive/atomic and necessitate all the facts (including the negative facts). Roughly speaking, positive/atomic facts are typically not able to necessitate negative facts because they can always be supplemented with additional positive/atomic facts. However, if the positive/atomic facts are plenitudinous in the way that **Island Universe Plenitude** describes, then this problem can be avoided.

One potential objection to this view is that, even if **Island Universe Plenitude** is true, it still seems that there could be additional numerically distinct but qualitatively indiscernible island universes. In response, I argue that **Island Universe Plenitude** should be supplemented with The Identity of Indiscernibles, according to which there cannot be numerically distinct entities that share all of their qualitative properties and relations. In other work, I have given independent reasons for such a principle as well (in particular, I've argued for a version of the Bundle Theory that entails The Identity of Indiscernibles).

Lastly, I argue that, not only is **Island Universe Plenitude** and The Identity of Indiscernibles *sufficient* to resolve the tension that we have been facing, but both views are also *necessary* to resolve the tension that we have been facing. I use this fact to argue that, if it is necessary that all fundamental facts are positive/atomic, then it is likewise necessary that **Island Universe Plenitude** and The Identity of Indiscernibles is true. I close the paper by addressing whether this last consequence can be used to provide an answer to the classic question "Why is there something rather than nothing?".

Connecting Language-Games: Diamond on Truth, Realism and The World

Barnaby Burleigh (Oxford, UK)

Abstract

Cora Diamond has argued that the real substantiality of truth lies in the connections we make between language-games. In this paper, I investigate what Diamond considers this real substantiality to be. Diamond's unfolding account of truth aims to steer clear of metaphysics, without reducing truth to a grammatical instrument. She argues that in the connections we make between language-games, it shows itself that there is a world. I suggest that Diamond is right to point to the connections between language-games as the key to understanding truth, but argue that Diamond's account leaves it unclear what her talk of there being a world is supposed to mean. I explore Diamond's work on realism as a sketch of how she might get to "the world". However, drawing on the work of Charles Taylor, I argue that there is a sense in which Diamond's realism is in fact dependent on the idea of a world, thereby problematizing Diamond's attempt to treat world-talk as a mere upshot of the connections we make between language-games. I sketch how the resulting dialectic might be resolved by exploring the sources of Diamond's realism. This paper, then, is an attempt to show that Diamond's work points us in the right direction if we are properly to understand truth, but that the path Diamond sketches out for unfolding the concept requires some adjusting, if it is to be successful.

1. Introduction: Dummett, Winch and Diamond

In "Unfolding Truth and Reading Wittgenstein", Diamond makes a distinction between substantial and non-substantial accounts of truth. She remarks that ordinarily, this terminology is used to distinguish 'deflationary' theories of truth from theories which take the concept of truth to have substantial content. Diamond believes that truth cannot be adequately *theorized*, but she still takes the substantial/non-substantial contrast to be helpful. By non-substantial accounts of truth, Diamond means accounts which treat truth as a mere grammatical instrument (akin for example to a pronoun). By substantial accounts, she means accounts which take truth to be more than merely grammatical (Diamond 2003: 25).

An unfolding account of truth is an account which refuses to analyse the concept of truth in terms of a general theory. The reason for eschewing theorizing is that Diamond believes, following Frege, that the substantiality of truth is something that comes out in "the normative constraints on judging,

where judging is taking to be true a content that can be judged true or false" (Diamond 2003: 25). In order to give an account of the concept of truth, we have to explicate these normative constraints.

The background against which Diamond formulates her position is Winch's critique of Dummettian realism. Dummett's realist holds that any true proposition is true in virtue of some fact. For example, the statement "Russell is in prison", uttered in 1918, is true in virtue of the same fact as "Russell was in prison in 1918", uttered today. Winch argues against the realist that it is a mistake to say that these two propositions are true in virtue of the same fact. The conditions which warranted the assertion of "Russell is in prison" in 1918 are very different from the conditions which warrant the assertion of "Russell was in prison in 1918" today. Today we have to rely on memory and historical documentation, whereas in 1918 we could have gone to visit Russell in his cell. Winch argues that if we are to understand truth, we need to pay attention to what actually allows us to call our propositions true. The appeal to a metaphysics of facts which simultaneously function as truth-makers for both present and past tense statements is meaningless, because it has no connection to how we actually establish truth in each language-game (Winch 1987: 38-46).

Diamond is sympathetic to Winch's critique of Dummettian realism, but worries that the idea that truth is established within separable language-games is a mistake. Winch's account of truth is substantial in that it does not reduce truth to a grammatical feature, but it lacks "real substantiality". According to Diamond, Winch's account obscures the connections that in fact exists between the two propositions about Russell, for example that we treat "Russell was *not* in prison in 1918" as *contradicting* the 1918 statement "Russell is in prison". Diamond suggests that such connections between language-games are quite central to the concept of truth. It is here that we encounter "the normative constraints on judging" which Diamond believes give the concept of truth its real substance.

Imagine two friends with a playful interest in astrology gossiping about other peoples' star-signs:

Friend A: "Peter and Mary will have a great marriage."

Friend B: "I don't think that's true, Peter's a Capricorn and Mary's a Sagittarius – they'll never agree on anything."

Consulting the zodiac shows Friend B's utterance to be true. Capricorn and Sagittarius are believed to be star-signs with low compatibility. "Peter and Mary probably won't have a great marriage" is true in virtue of the fact that Peter is a Capricorn and Mary a Sagittarius. This is the correct elucidation of truth in the zodiac-interpreting language-game. While there is nothing wrong with this analysis, as far as it goes, Diamond suggests that if we are properly to unfold truth in a case like this, we have to go further. Because of course it is *not* true that Peter and Mary are less likely to have a happy marriage due to their star-signs. Even though there is a context in which Friend B was right to assert what he did, and even entitled to claim that what he was saying is true, in fact, what he said is false. The upshot of this sort of case is that "the real substantiality of our conception of truth is tied to the ways in which we take the connections of different language-games". (Diamond 2003: 42). I want to explore what this real substantiality might be thought to be.

2. A World that Shows itself

In "Unfolding Truth", Diamond writes the following:

[J]ust as the game in which we make statements that are about the past has tense internal to it, and hence the statements made in it are logically related to statements not in the past tense, the activity of using which is quite different, so some of our language-games have, internal to them, that what is said in them is true, if it is true, in virtue of circumstances that form a world (as statements in the past tense and in the present are true, if true, in virtue of circumstances that together form a temporal order). To say "There is a world" would be to try to say what shows itself in the compatibility-relations and incompatibility-relations of things said in various games, what shows itself in the various truth-games (as we might call them) not being self-contained, as the game of making past-tense statements is not self-contained. (Diamond 2003: 49)

The connections in which Diamond is interested are supposed to render talk of "circumstances that form a world" meaningful, without relying on the

metaphysics of Dummettian realism. Her account, as opposed to Winch's, still manages to offer real substantiality, precisely by leaving room for this sort of world-talk. How is this supposed to work?

Diamond wants to suggest that the fact that we take our statements *about* the past to be capable of contradicting statements that were made *in* the past shows us something about tense. Our life is tensed, or, as she puts it, "there is a temporal order". Diamond does not mean for this to involve a metaphysical claim. What she means to bring to our attention is just the fact that we make these connections between language-games. To say that there is a temporal order says no more than what is already there in the relations that obtain between differently tensed statements.

This idea about the relations between differently tensed statements is the basis for the analogy which Diamond makes in the above quotation. She thinks that the connections we make between different language-games which involve statements that can be construed as truth-claims show that "there is a world". Diamond seems to be right to call our attention to these connections. The astrology example is a case in point, as are tensed statements, or the many cases where scientific study has led to the reassessment of beliefs that were taken-for-true on grounds that are no longer acceptable to the modern scientist. The question is to what extent calling attention to these connections gives us an account of truth which has real substantiality without being metaphysical. How do these connections amount to the idea that the claims made in the language-games which are being so connected are "true, if they are true, in virtue of circumstances that form a world"? Diamond seems to want to say, by analogy with the case of temporality, that this idea is simply an upshot of the connections we do make. As she puts it: "To say 'There is a world' would be to try to say what shows itself in the compatibility-relations and incompatibility-relations of things said in various games" (Diamond 2003: 49). Diamond's argument in a nutshell seems to be this: *pay close attention to the connections we make between different language-games and you will see how the concept of truth functions. Then, if you tried to say how it functions, you would probably be tempted to speak of "circumstances that form a world". You can in fact speak in this way without embracing the metaphysics of Dummettian realism by acknowledging that all we can do to give meaning to this phrase is to point to the connections we make between different language-games. Hence we*

get to the idea that the concept of truth has a real substance to it, without doing any metaphysics.

3. Truth and Realism

The task then is to explain how Diamond's talk of showing that "there is a world" links to the connections we make between language-games. The most promising strategy for understanding the relation between truth and world which Diamond sketches in "Unfolding Truth" seems to be to investigate what she says elsewhere about realism. The connections we make across language-games when assessing whether propositions are true are made via the question whether what is taken to be true in a particular language-game is *really* true. In "Realism and the Realistic Spirit", Diamond calls this question an application of elementary realism (Diamond 1991, 53-55). Diamond's sketch of what this elementary realism amounts to involves three strands:

1. Facing the facts.

a) not refusing to look at the facts.

b) not assuming that the facts are a certain way just because you think that they *ought* to be.

2. Not buying into magic, myth, fantasy, superstition.

3. Paying attention to causation, i.e., being aware of how things actually work, making sure your thoughts and plans cohere. (Diamond 1991: 39-41)

These characteristics of realism are clearly important when it comes to making the connections between language-games which characterize truth. Within a given language-game, there can be magic, myth and fantasy. Within a given language-game, causation need not matter, nor need facts. The point of Diamond's account of truth is that the truth-claims made within such language-games can come under the scrutiny of elementary realism, as evidenced by the astrology case above.

How might the idea that "there is a world" be thought to show itself in the connections we make by bringing to bear elementary realism across language-games? Is the answer just that people who buy into this realism are inclined to

say that there is a world? This is all that Diamond seems to give us, although the "showing"-terminology hints that she thinks that something more can be said. If you claim that something shows itself, it ought to be possible to say how. The problem is that none of the connections we make, taken as examples in isolation, will be enough to make the Diamondian point. What Diamond is saying is that it is in the whole practice of connecting language-games that it shows itself that there is a world. But if we try to say something about this whole practice of making connections, it is extremely difficult not to slip back into metaphysical talk. The idea seems to be that the fact that the constraints of realism can be brought to bear across different language-games, quite independently of their respective conventions, shows that there is something independent which plays a role in determining truth. The open-endedness of truth is a reflection of the fact that it is not up to us to determine what counts as true. But how is saying this any different from Dummettian realism? Because surely, once we have said this much, it is hardly a further step to say that truth is determined by the world?

The answer seems to be that Diamond is perfectly happy to say that truth is determined by the world, so long as this kind of talk is taken as a mere upshot of the connections we make in language, rather than being understood as a foundation for such connections. As something that is *shown*, this world-talk means something quite different in Diamond than it does in the mouth of Dummett's realist. All Diamond seems to be saying is that *if* we pay proper attention to the connections between language-games, we will end up with a conception of truth which might prompt us to reach for world-talk. The problem is that at this level of generality, Diamond's idea that there is a world is a mere pointer which we cannot properly flesh-out without actually unfolding the concept of truth. You do the unfolding, and, if you do it properly, Diamond predicts, you will see why someone might want to say that there is a world. That we are tempted to say this is a symptom of the real substantiality of truth, rather than giving the concept of truth its actual content. That content resides purely in the connections we make between language-games, the concept's normative relations.

What is concerning about this solution is that it seems to want to connect world-talk with the real substantiality of truth, without giving an account as to how the connection actually works. What does it mean to say that what is

really true "is true in virtue of circumstances that form a world"? Explanation is deferred as something that will become redundant as the concept of truth is unfolded. Once it shows itself in the unfolding process that there is a world, you will understand all that there is to understand in this phrase. This 1) leaves it obscure what Diamond's world-talk is supposed to amount to, and 2) relegates world-talk to the status of a *product* of the connections we in fact make between language-games. We could just accept both of these as natural features of an unfolding account. However, I want to suggest that there are reasons not to be content with this suggestion and to take investigation in a different direction.

4. Realism and the World

If we ask where the elementary realism Diamond sketches in "Realism and the Realistic Spirit" comes from, it becomes questionable whether world-talk should indeed be thought of merely as an upshot of an established practice. The connection between elementary realism and truth, which we have been trying to unfold, is historically contingent. People were not always unwilling to call magical, mythical or fantastical explanations true, and our conception of what makes for a plausible causal story is certainly much changed since mediaeval or ancient times. The crucial question then is why the connections of elementary realism have the force they do. Why does elementary realism win out when we bring it to bear on astrology?

Charles Taylor has offered an interesting genealogy of what he calls "the naturalism of disengaged reason", which explores the historical origins of the kind of elementary realism which Diamond articulates in her work on truth (Taylor 1989: 495). Taylor's work suggests that any plausible genealogy of elementary realism must involve the idea that there is a world as among the forces that bring about our strong commitment to this realism (Taylor 1989: 337-351). The process of disenchantment, which furnishes elementary realism with its opposition to myth and fantasy, according to Taylor, is dependent on the emergence of the concept of a "buffered"-self, i.e., a self which is not vulnerable to spirits, no longer porous. It is shielded from a "world", which lies outside, as something which we encounter (Taylor 2007: 30-43). If this is right, then the force of elementary realism comes in part from the very conception

of the world which Diamond wants to treat as its product (namely as something which is shown). This suggests that there is after all a sense in which the connections we make depend on the idea that there are circumstances that form a world, rather than vice versa.

How might this dialectic be further developed?

1) Diamond might argue that to treat the idea of a world as more than an upshot from connections we make, i.e., to treat it as more than something that "shows itself", is always metaphysical. However, showing Taylor's idea of "the world" to be metaphysically confused would not be enough to falsify his historical analysis. A metaphysically confused idea can still play an important role in establishing and promoting a new way of thinking. Bearing in mind the way in which the likes of Richard Dawkins wield elementary realism as a tool for exposing all religious and much moral thought as fantasy suggests that there is reason to suspect that even contemporary elementary realism may be bound up with metaphysical prejudice. And even if Diamond can defend the idea that the connections we now make between language-games do not depend on metaphysical confusions, confidence in our current practice may still be diminished, if metaphysically confused ideas turned out to be among the historical sources of elementary realism.

2) Diamond might be brought around to the view that there is a way of giving meaning to world-talk which avoids the metaphysics of Dummettian realism, but does not make the idea that there is a world something that *shows* itself. Dreyfus and Taylor's *Retrieving Realism* could be read as an attempt to give such an account (Dreyfus and Taylor, 2015), which, if successful, might render the complex relation between elementary realism and world-talk sketched in Taylor's genealogy acceptable to a reader of Diamond.

These suggested developments take seriously Taylor's idea that the force that elementary realism has for us today is best understood by paying attention to its historical origins. Given the connection between realism and disenchantment (a concept which has a temporal dimension very much built into it), this seems a plausible approach. While Diamond's resolute commitment to making sense of the real substantiality of truth without falling into metaphysical confusion is admirable and should not be abandoned, treating

this real substantiality merely as something that shows itself in the connections we make between language-games seems an insufficient characterization. It misses the complex role that ideas of the world have played in giving the connections of elementary realism the central place they currently have in our language. What the real substantiality of truth consists in remains to be shown by a careful analysis of the sources which give the connections between language-games their force.

References

Diamond, Cora (1991) "Realism and the Realistic Spirit", in her: *The Realistic Spirit*, Cambridge MA: MIT Press.

Diamond, Cora (2003) "Unfolding Truth and Reading Wittgenstein" *SATS – Northern European Journal of Philosophy*, Vol. 4, No. 1, 24-58.

Dreyfus, Hubert; Taylor, Charles (2015) *Retrieving Realism*, Cambridge MA: HUP.

Taylor, Charles (1989) *Sources of the Self*, Cambridge MA: HUP.

Taylor, Charles (2007) *A Secular Age*, Cambridge MA: HUP.

Winch, Peter (1987) "Im Anfang war die Tat", in his: *Trying to Make Sense*, Oxford: Blackwell.

Acknowledgement

Many thanks to Theo Hickfang for helpful comments on an earlier draft of this paper.

Logic as Metaphysics in Wittgenstein

Bangrui Chen (Chicago, USA)

Abstract

This paper offers a reading of Wittgenstein's conception of logic. The stance I intend to defend is the one I call "logic as metaphysics". The thesis has its roots in Aristotle, and I think it can be applied to both the early and the late Wittgenstein. The investigation serves two primary aims: firstly, I elucidate how this perspective problematizes conventional notions of logic, and secondly, by doing so, I hope to highlight important continuities in Wittgenstein's philosophical trajectory. In section 2, I provide the sketch of the view I call "additive theory of logic" as an opponent to Wittgenstein's conceptions of logic. In section 3, I provide and defend my reading of the early Wittgenstein with then lenses borrowed from the observation of logic as metaphysics. I also argue how some readings on Wittgenstein fall into the additive theory and risk deviating from Wittgenstein's main concerns. Section 4 adopts the same approach to the late Wittgenstein. I will conclude in Section 5 by underscoring the philosophical merits of this interpretive framework.

"Now it is becoming clear why I thought that thinking and language were the same." Wittgenstein, *Notebooks*, 1914-1916

"Like everything metaphysical the harmony between thought and reality is to be found in the grammar of the language." Wittgenstein, *Philosophical Grammar*

1. Introduction

In this paper, I aim to offer a reading of Wittgenstein's conception of logic. The stance I intend to defend is what I call "logic as metaphysics", which I think can be applied to both the early and the late Wittgenstein.

Contemporary philosophy often treats logic as a discipline that studies the formal principles of thinking. In ancient times, however, Aristotle suggests a broader scope of logic when he first enunciates the idea that logic is metaphysics. He defines nature as the form specified by logos (*Physics* 193a30-31), implying that logic concerns not only the laws of our thought, but also their relation to the world. This is reminiscent of the doctrines specified by Heraclitus (Fragment 1) and Parmenides (Fragment B3), that all beings are in accordance with logos and that it is the same thing to think and to be. What they suggest in these sentences, presumably, is that there is one and the same logic underlying and ruling both nature (*phusis*) and the mind (*nous*).

I argue that this Aristotelian theme reappears in Wittgenstein's work, and it is such conception of logic that we seek to explore in this paper. By elucidating this perspective, I also hope to underscore some key continuities in Wittgenstein's thought. As articulated in the *Tractatus (TLP)*, the world, that which is, must be logical; in the *Philosophical Investigations (PI)*, he maintains that notions such as rule-following process will not abolish a certain kind of logic. While Wittgenstein changes his mind on many things in different periods, he consistently adheres to this conception of logic throughout his life.

I will proceed in stages. Section 2 outlines what I term the “additive theory of logic”, which serves as an opposing viewpoint to Wittgenstein's perspective. In section 3, I provide and defend my reading of the early Wittgenstein, demonstrating the way in which he expresses the idea of logic as metaphysics. Section 4 extends this approach to the late Wittgenstein. I also expound on how some readings on Wittgenstein fall into the additive theory and risk deviating from Wittgenstein's main concerns. I conclude this interpretative framework in section 5 and highlight its philosophical merits.

2. Two Layers of Additive Theory of Logic

Taking logic as a theory of the form of pure thinking, and metaphysics (particularly after Kant) as a science of being grasped by the knowing subject, the expression “logic as metaphysics” denotes an identification of these two branches of philosophy. In order to better characterize this stance, I will briefly discuss its opponent view, which I name as *Additive Theory of Logic (AT)*.

AT does not accept the Aristotelian identification of logic and metaphysics. It hence does not accept a presupposition of any original identity of thinking and being. Instead, a supporter of AT will propose the following point: logic exists only in one of the two realms (either thinking or being), and its occasional extension to another realm is merely an additive process. This extension is never fundamental nor necessary, or, at least, this process cannot be proved to be fundamental or necessary.

There can be two layers of AT, namely the *Onto/Logo Dualism* and the *Psycho/Logo Dualism*. The first layer concerns beings qua objects and facts, the

existence and the givenness of which are independent of any logical functioning of the mind. As a result, logic is said to be added to the objective side. The second layer concerns true sentences or propositions, which have objectively logical validity prior to and independent of the mind's recognition and affirmation of it. As a result, logic is added to the subjective side. To better illustrate the point, I will now turn to some of Frege's and Russell's doctrines as representatives of AT.

Russell's early version of logical atomism, taken from the ontological level, can be seen as an example of the first layer of AT. The tenet of logical atomism consists in the belief that the world is composed of absolute simples (e.g., entities and the qualities they exhibit) that are devoid of any structure, including logical structure, among them. If so, then the world is ultimately reducible to simple objects that "do not presuppose complexes" and that "have a kind of reality not belonging to anything else" (Russell, *Analytic Realism*, 94; Russell, *The Philosophy of Logical Atomism*, 270). Therefore, since all simple objects and the atomic facts made up of them are logically independent of one another, it is correct to infer that the logical form between any two objects or any two facts cannot be self-generated but must be additive from outside.

I take Frege's notion of logic as an example of the second layer of AT. According to Frege, the subject matter of logic is not within the mind itself, but in some abstract orders holding among propositions. Logic studies not the mind (as psychologism proposes), but the truth. On the other hand, these orders of propositions must prescribe and reflect how we ought to think and how we actually think. It sounds puzzling, however, that something whose subject matter is not the mind can prescribe what our mind should think. Frege addresses this problem by introducing the judgmental stroke (or, the assertion sign, "┐") as the sign of the assertoric force. The reason why Frege needs such a sign is his conviction that we are able to infer logical truth. But inferences themselves are not truth; the principles of inference are not themselves logical laws of truth. Instead, an inference is an act according to rational permissions to access logic. Similarly, Frege later says that the assertion is the act of our declaration or acknowledgment of the truth of thought in a sentence. Yet the assertion sign, which marks an act of the subject,

stands external to logical propositions. Logic standing on its own hence becomes unassertive and forceless to our mind (see Frege, *The Basic Laws of Arithmetic* 37-39; *Thought* 62-63).

Having outlined the two layers of AT instantiated by Frege and Russell, I will now examine how the early Wittgenstein's conception of logic diverges from both layers of AT.

3. The Early Wittgenstein against the Additive Theory of Logic

Some interpreters endorse a view that the early Wittgenstein, following Frege and Russell, treats logic in a *merely* formal or contentless manner (see, e.g., A. Maslow 11-2; W. Goldfarb 172; O. Kuusela 13). I do not want to deny all features Wittgenstein assigns to logic that seem formal, particularly not those related to the atomist thesis concerning logically independent propositions. But the danger lying behind these formalist readings is that they blur the distinction between early Wittgenstein's stance and the universalist conception of logic, the latter of which might easily slide into AT, as demonstrated by the cases of Frege and Russell in the previous section. In this section, drawing on sources from his *Notes on Logic (NB)* and *Tractatus Logical-Philosophicus (TLP)*, I argue that the early Wittgenstein distances himself from both layers of AT.

First, regarding the onto/logic dualism, unlike Russellian ontological atomism, Wittgenstein argues that objects as contents are innately connected with their form, even though he occasionally expresses the idea such as "objects are simple and cannot be composite". An object can still be said to be complex in a sense that it is a combination of matter and form. A simple object is an object obtaining a first-order determination substantially, not an object devoid of any *potential* determination from the world's logical form (TLP 2.021-2.025; B. Allen 219-220).

The logical objects in *Tractatus* language are not things in the ordinary sense, for these objects are not empirical – if logic is transcendental (and Wittgenstein thinks that it is), so are the objects. Logic and objects are interdependent to each other, precisely because the very notion of form presupposes its content and is also presupposed by its content (TLP

2.0211-2.00212; Anscombe 165-6). On the one hand, logic cannot merely be an additive form applied to a pure mass because it cannot be prior to the question “what” (TLP 5.552). On the other hand, logic cannot pre-exist independent of the world; if that were the case, given that there is a world, logic would have little to do with *this* world (TLP 5.5521). As an analogy, while Kepler’s laws predetermine the positions of every possible planet, it makes no sense to say that these laws exist prior to the existence of any planet. Similarly, the logical laws prescribe the world’s way of being but cannot be independent of that which exists in order to be applicable to objects in the world. Otherwise, if we grant that logic can exist alone, it remains unexplainable how logic can be applied to *the world*.

Second, regarding the psycho/logo dualism, Wittgenstein would respond that first-personal acts to assert/judge/think a logical truth is the same as that logical truth itself. That is, in any case, the same “p” stands in the proposition sign “p” and the “X thinks/judges p”.

By Frege’s assertion sign, we have seen that “ $\vdash p$ ” and “p” stand for two very different things. The latter is the container of a thought, while the former of them concerns our mind’s activity in regard to this thought. The idea that the assertion symbol is completely outside of the content of a proposition indicates that the unity of propositions themselves is independent of the subject’s consciousness of that unity. By contrast, Wittgenstein takes the Aristotelian stance that every proposition must contain an assertoric force within itself (see I. Kimhi 47). But the assertion is also psychological, since the proposition “p” displays a possible act of our mind; the contents of both the proposition and my mind’s act are identical (NB 93ff.). It is in this sense that Wittgenstein holds that thought cannot be illogical (TLP 3.03, 5.4731). According to this, the content of any good proposition “p” and that of my consciousness of this proposition, namely “I think ‘p’”, are identical, because my consciousness of such propositions must be both an act of my consciousness and contained within my consciousness. On the other hand, thoughts like “p&p” and “p→q, p, hence -q” will not be possible acts of our consciousness; strictly speaking, expressions like “p&p” are not propositions at all.

4. The Late Wittgenstein against the Additive Theory of Logic

Common readings of the late Wittgenstein often view themes such as grammatical statements and rule-following considerations in *Philosophical Investigations* (PI) as a departure from his early account of logic. People disagree with where and how Wittgenstein expresses dissatisfaction with the doctrines of the *Tractatus*. But this is not my focus here. The thesis I will argue for in this section is this: though the late Wittgenstein gave up many of his earlier conceptions of logic, such as the switch from logical atomism to logical holism (D. Stern 98ff.) or the subsuming of the universal and calculus-based logic as an element of a broader logical methodology (Kuusela 40ff.), his adherence to the identification of logic and metaphysics remains unchanged, and he is also consistent in against for the two layers of AT.

Once again, this move is not to reject many philosophical insights observed in PI by commentators. Instead, I intend to show, given such discontinuities between the two Wittgensteins, how we can make sense of some claims on logic we encounter in PI, both exegetically and philosophically. To begin with, an initial point of evidence can be found in §242, where Wittgenstein notably speaks to his interlocutor that his rule-following treatment of language “seems to abolish logic, but does not do so”. I propose that the pivotal point here is to determine what kind of logic he is talking about in this sentence. I propose that the term “logic” in this sentence should not be taken in a general sense, but be properly understood only in a metaphysical sense, indicating an implicit line of continuity between TLP and PI.

Consider the first layer of AT. Let me take Stern’s and Kuusela’s readings as examples, to which I have little objection. But I think the development of Wittgenstein’s conception of logic in these aspects do not alter or devalue his previous thoughts that logic is not additive from a subjective or quasi-subjective realm to the objective world.

In §142, Wittgenstein urges readers to think of a case in which the practice of selling cheese by weight loses its point. As he states, “if things were quite different from what they actually are – [...] if rule became exception, and exception rule; or of both became phenomena of roughly equal frequency – our normal language-game would thereby lose their point.” Prior to that section, Wittgenstein reiterates that the meaning of a word cannot be

identified as an isolated thing that is independent of the context or use, and given a certain word, we can always come up with an imagination within which the meaning of that word is misunderstood. In daily life, the conformity between a word and its application is in determination since we are often in normal cases, cases in which our mindset has previously been taught to act in a certain way in this or that situation. Thus, Wittgenstein notes in the same section that to explain the significance of a concept, we need to consider some “extremely general facts of nature”, facts “as are hardly ever mentioned because of their great generality”. These general facts of nature, not only correspond to the objective features of the world’s appearance, but also innately correspond to the formation of our mindset.

A rule should tell me the way of its application, just like a measurement should tell me the weight of an item. I think the metaphysical sense of logic in Wittgenstein consists in this: there has been a tacit agreement between the way the world we live in is and the nature of our practice in grasping the manner of the world’s functions, which is prior to our rule-following activities and our recognition of our ability to follow rules. Thus speaking, the logical structure, no matter whether it is the rule, or the grammar, or what have you, is still not a subjective product external to and imposed on the world.

Hence, my construal of PI’s conception of logic differs from the *transcendental* or *idealist* reading which sees the structure of our world as predetermined by some of our subjective settings (P.M.S. Hacker 179; J. Lear 223; T. Nagel 105). This is because in PI there is not a substantive boundary between what is empirical and what is transcendental (for the certainty of our practices is not lying outside the empirical world). However, there is also a significant distance between my stance and the *naturalized* reading of PI which states that logic is merely contingent, if by contingent one means the validity of logic/grammar is possible merely by chance (N. Malcolm 18-9; P. Maddy 86-7; compare this idea with PI §497). Logic is not contingent if we restrict our sight to every single possible world in itself. In any possible world, say, for instance, the world of nut-calculator in *Remarks on the Foundations of Mathematics* (Part I §137) or the Martian in PI §139, the interrelation and interaction between the nature of actors and the nature of the world is fixed to a certain degree.

As for the second layer of AT, it would suffice to refer back to PI §242 again. Wittgenstein emphasizes here that to enable communication by means of language, not only *agreement in definitions* but also *agreement in judgements* is required. This latter agreement, is not determined by the concepts of true and false, and hence, by our application of the calculus of truth functions to our language (PI §136). Rather, when sticking to the grammar-in-practice thesis, we reject the picture in which the use of a predicate in a proposition is logically external to the psychological aspect of our mind (for elaborations on this point, see Kimhi 51). Logic hence is immanently located in my consciously making any judgement. Wittgenstein is in no way denying the existence of thought and its logical form, but he argues that it is impossible to separate the “essence of thinking” (i.e., logic) from the process of my actual thinking, that is, to separate it from a concrete realm in which I can perform and use it under certain descriptions of my judging acts (PI §97). What Wittgenstein implies by saying that the “measuring” is partly determined by a certain constancy in the results of measurement, is this: the methods of measurement (denoting logic) are, to some extent, existing in the constancy in speakers’ agreement in their real acts of making judgements. As in *Tractatus*, for late Wittgenstein it makes no sense to see logic as something detached from the functioning of the mind, only being applicable to and determining our actual thoughts from an independent domain, as Frege’s view seems to imply. The real difference between his early and later positions lies in the number of logics: while in *Tractatus* there seems to be a pure, crystalized logic, later he detects the multiplicity of grammatical systems, the number of which is equal to the number of all possible language games.

5. Conclusion

The problem with AT lies in its oscillation of logic’s position between subjective and objective realms. In its first layer, logic is subjective in the sense that it must be found as an aspect of the functioning of our thoughts and must exist outside of the beings themselves. In the second layer, logic is objective in the sense that the rules of thoughts must not be equated with all activities of our mind; otherwise, either there would be no criterion for distinguishing false judgments from correct ones, or anything I think would become logical (even something like “p&-p”).

Attempts to resolve this dilemma, such as Frege’s assignment of “objective thought” to a third realm, only seem to complicate the issue for Wittgenstein. This is because, as he argues, applying logic to both realms – objective and subjective – creates additional problematic gaps. We may then ask, what philosophical merits we can draw from his objection to AT and his endorsement of the doctrine of logic as metaphysics? I conclude as follows:

First, inspired by Aristotelian hylomorphism, the logical form and logical matter are only reflectively abstractable/distinguishable, but not ontologically extractable/separable. As an analogy, we can think of a music passage’s pitch and timbre, which are different but cannot be separated. Second, any logical truth must stand as the readiness (potentiality) of its completion (actuality) i.e., my grasping it in my consciousness. A proposition standing alone is the same as my ascription of it in thinking. As an analogy, we can think of a living creature, in which its actions of eating and digesting cannot be taken as two parts, but as different stages of the same process.

Viewed from this monist perspective, the seemingly separated parts taken by dualists are actually identical and different at the same time. Contrary to Frege’s terms, logic does not exist in an independent third realm. Instead, Wittgenstein suggests that logic is not detachable from both the realms of thinking and that of being. If both branches of my interpretation of Wittgenstein’s objections to AT are correct, then it is also correct to say that Wittgenstein, both early and late, strives not to separate thinking and being, but to bring them into a possible unification.

References

- Allen, Barry (2004). “Wittgenstein’s Onto-Logic”, in: Luitzen Egbertus and Jan Brouwer, et al (eds.), *Post-analytic Tractatus*. Aldershot, Hants, England; Burlington, VT: Ashgate.
- Anscombe, G. E. M. (1965) *An Introduction to Wittgenstein's Tractatus*. 2d ed., rev. New York: Harper & Row.
- Frege, Gottlob, and Montgomery Furth (1964) *The Basic Laws of Arithmetic: Exposition of the System*. Berkeley: University of California Press.
- Frege, Gottlob, and P. T. Geach (1977). *Logical Investigations*. New Haven: Yale University Press.

- Goldfarb, Warren (2018). "Wittgenstein against Logicism", in: Erich H. Reck (ed.), *Logic, Philosophy of Mathematics, and their History: Essays in Honor of W. W. Tait*. College Publications.
- Griffin, James Patrick (1997) *Wittgenstein's Logical Atomism*. Bristol: Thoemmes Press.
- Hacker, P. M. S. (1986) *Insight and Illusion: Themes in the Philosophy of Wittgenstein*. Rev. ed. Oxford [Oxfordshire]: New York: Clarendon Press; Oxford University Press.
- Kimhi, Irad (2018) *Thinking and Being*. Cambridge, Massachusetts: Harvard University Press.
- Kuusela, Oskari (2022). *Wittgenstein on Logical and Philosophical Method*. Cambridge: Cambridge University Press.
- Lear, Jonathan & Stroud, Barry (1984). "The Disappearing 'We'", *Aristotelian Society Supplementary Volume* 58 (1), 219 - 258.
- Maddy, Penelope (2014) *The Logical Must: Wittgenstein on Logic*. New York, NY: Oxford University Press.
- Malcolm, Norman (1986) *Nothing Is Hidden: Wittgenstein's Criticism of His Early Thought*. Oxford, OX, UK; New York, NY, USA: Blackwell.
- Maslow, Alexander (1997) *A Study in Wittgenstein's Tractatus*. Bristol, England: Thoemmes Press.
- Nagel, Thomas (1989) *The View from Nowhere*. New York: Oxford: Oxford University Press.
- Russell, B. (2003) "Analytic Realism", in: *Russell on Metaphysics*, ed. S. Mumford. London: Routledge.
- Russell, B. (1956) "The Philosophy of Logical Atomism", in his: *Logic and Knowledge*, ed. R.C. Marsh. London: Allen & Unwin.
- Stern, David G. (1995) *Wittgenstein on Mind and Language*. New York: Oxford University Press.
- Wittgenstein, Ludwig (1961) *Notebooks, 1914-1916*. Oxford: Blackwell.
- Wittgenstein, Ludwig (1974) *Tractatus Logico-philosophicus*. [Revised ed.]. London: Routledge and Kegan Paul.
- Wittgenstein, Ludwig, G. H. von Wright, Rush Rhees, and G. E. M. Anscombe (1978) *Remarks on the Foundations of Mathematics*. Rev. ed. Cambridge, Mass.: MIT Press.
- Wittgenstein, Ludwig, P. M. S. Hacker, and Joachim Schulte (2010) *Philosophical Investigations*. 4th ed. Hoboken: Wiley.

Sortal Terms and the Puzzle of Coincidence

Zhengguan Chen (Bern, Switzerland)

Abstract

Is an ordinary object, say, a statue, identical with its underlying matter, a piece of copper, which occupies the same region as the statue, yet has properties not possessed its coincident? This puzzle (typically under the name ‘the puzzle of coincidence’, has been discussed in a variety of dimensions: whether the arguments for non-identity are invalid, unsound, or not really by the so-called Leibniz’s Law; whether the intuition about the impossibility of place-sharing is legitimate; and whether such intuition and the arguments for non-identity are really incompatible. In this paper, I aim to show that the puzzle of coincidence arises mostly due to our linguistic practice of employing sortal terms, that is, the terms that denote kinds. For such purpose, a set of features of sortal terms will be laid out, and arguments for the non-identity of coinciding objects will be analysed in these terms. This view, as I see it, would imply that the truth conditions of our ordinary discourse do not extensively run afoul of the underlying nature of the world.

1. Introduction

Sortals, such as ‘tiger’, ‘water’ and ‘statue’, in contrast with ‘big’, ‘tasteless’, or ‘well-made’, are taken to denote kinds. As will be made clearer, I conceive of sortals to be those terms that can be used to answer the ‘what is it?’ question: suppose there is something on the table. When I ask my interlocutor what that is, ‘A book.’ would be a satisfactory answer, while ‘Something brown and white.’ or ‘Something weighing 0.5 kg.’ would not, plausibly, tell me what that thing is. (For my present purpose, I will not engage in the analysis of the relation between kinds and properties in this paper. This issue is, however, significant for bridging the fundamental reality and the ordinary language discourse that I aim to examine. Chapter 2 will be devoted to this topic.) ‘Book’, as the test case shows, is thus a sortal term, while ‘something brown and white’ and ‘something weighing 0,5 kg’ are not. Or so I will argue.

Why should we bother with sortals? As I see it, one of the most important reasons to be concerned with sortal terms is that sortal terms are at the very intersection of the study of ordinary language discourse and the study of the fundamental structure of reality. To begin with, we customarily refer to both ordinary objects and physical, chemical, or biological objects investigated in particular sciences in sortal terms to achieve our communicative purposes. It is unintelligible, if not impossible at all, to baptize every individual for the mere purpose of referring to them. Since sortals are terms that denote kinds as

introduced above, our linguistic practice implies that things can be grouped under different kinds. A question then follows when we have sortal terms at our disposal: are the ways we refer to a group of objects *natural* ways of classifying things? Or are the ways random, or no more than gerrymandering? It seems the most natural thing to group two copies of Sider's *Writing the Book of the World* together and to leave a cow, for instance, out of this group. But things might become more complex once we put the philosopher's hat on. Furthermore, it is also contentious whether there are such entities as natural kinds, that is, whether the notion of natural kind belongs to our categorial ontology, and how do natural kinds interact with other entities in our ontology, for instance with properties, objects, and events (processes).

Secondly, sortal terms are widely employed in a variety of metaphysical puzzles including the problems of coincidence, of composition, of the many, of ontological dependence. In fact, as I see it, these metaphysical puzzles arise mostly due to our employment of sortal terms in the relevant arguments. In the remainder of this chapter, I will seek to display how the metaphysical puzzles are triggered by our usage of sortal terms.

I am, of course, not the first to examine the notion of sortal terms. Burke (1992, 1994) examines the relation between objects, kinds, sortal terms and persistence conditions, Lowe (2006, 2009) offers a more comprehensive study of sortal terms based on what he labelled 'four-category ontology', that is, a realist theory proposing four most basic and irreducible concepts of ontology, i. e. objects, kinds, properties, and modes. Mooney (2023, 2024) defends what he calls phasalism about sortal properties, proposing that sortal properties ordinary objects instantiate are phase sortal properties, the properties objects can start or stop to instantiate without beginning or ceasing to exist.

However, Burke's doctrine of dominant sortal terms is, as I argued in a previous paper, incomplete in that it does not place any restriction on the kinds that objects belong to which undermines the argumentative force of his doctrine. Lowe's theory, on the one hand, is comprehensive and clears ways to employ the ontological notions that we would like to have, on the other hand, however, it does not address the issue of the relation between the fundamental structure of reality and our ordinary discourse, e. g. how the truths of our ordinary discourse about existence, composition, and coincidence can still be

maintained in the face of objections from particular sciences that acknowledge the existence of only the most basic simples. Mooney's theory, as he concedes, offers merely an alternative solution to the puzzle of coincidence that is no better off than the rival strategies.

Therefore, standing on the shoulders of the precursors, I aim to propose a theory of sortals which would connect to the categorial study of ontology, metaphysical puzzles and the relation among the ordinary discourse, ordinary objects, and the underlying nature of the world, but which tries to pinpoint and to address the drawbacks and weaknesses of their theories, by investigating, most importantly, the naturalness aspect and the essence aspect of sortal terms. In this paper, I will motivate the study of sortal terms by considering the puzzle of coincidence, which, as I see it, is not genuinely a metaphysical problem in the sense in which it pertains to the underlying reality of the world, but rather arises because of our linguistic practice of employing sortal terms.

2. Features of 'sortals'

Let me start by displaying the available conceptions and the reason why I prefer one of them for my argumentative purpose. A variety of conceptions of 'sortal' has been proposed since its first usage by Locke (2008 [1690]). Generally, four aspects have been appealed to in conceiving of 'sortal'. The four corresponding conceptions are listed below:

- (1) The essentialist conception: a sortal tells us what the essence of an item is.
- (2) The counting conception: a sortal tells us the criterion for counting items of a kind.
- (3) The persistence conception: a sortal tells us under what conditions items persist.
- (4) The individuation conception: a sortal tells us the answer to the question 'What is it?'

As I see it, the notion of essence is itself not univocal; we would not want to determine the extensions of a notion based on another ambiguous notion. Therefore, I will leave out this conception in the first place.

The subsequent three conceptions seem to have rather unambiguous intensions. Certain terms seem to be good candidates that satisfy all these conceptions: ‘table’, ‘dog’, ‘apple’, ‘car’, etc. Nevertheless, there are sortal expressions that may meet one conception but not the other. Phase sortals ‘kitten’, for instance, permit counting: we may count how many kittens are there in the room; however, it does not provide the conditions under which the instances begin and cease to exist: a kitten can become adult without ceasing to exist. Therefore, phase sortals like ‘kitten’ satisfy conception (2) and (4) but not conception (3). Sortals such as ‘wave’, ‘volume of fluid’, or ‘garden’, as Wiggins (1980) suggests, can be individuated but are not countable. If he is right, then some sortals satisfy the conception (4) but not conception (2). With these different instances considered, I will adopt the individuation conception of sortals in this paper to include as many proper sortals available to us as possible, so that our analysis is not restricted by conception in the first place.

After offering my conception of sortals, I would like to display what I take to be the most uncontroversial features of sortal terms. These features would constitute a basis for my following discussion.

To begin with, the notion of sortal terms is introduced as the name for kinds. In other words, sortals are terms that denote kinds. No matter whether only natural kinds are part of reality or whether kinds are reducible to properties or other entities, it is undeniable that we have sortal terms at our disposal that seem to suggest there being entities of kinds. I will not deal with the realism about kinds, categorial ontology and criteria for naturalness now. They will be the subject of the ensuing chapters. What I want to maintain at the phase is simply that sortals are utilized to refer to kinds and not properties such as ‘red’, ‘sweet’, ‘wooden’ or ‘alive’.

Besides, associated with any sortal term, there is typically a series of predicates that one can naturally use to describe instances of the sortal term (whether truly or falsely) (Cf. Burke (1992, 1994); Fine (2003); Almotahari (2014); Sattig

(2015).). This series of predicates we call the sphere of discourse of the sortal term. Chairs, for instance, typically serve a particular function and can be meaningfully said to be ergonomic and well-designed; ‘ergonomic’ and ‘well-designed’ belong to the sphere of discourse of the sortal ‘chair’. On the contrary, pieces of wood do not typically serve such a function and thus cannot be meaningfully said to be ergonomic and well-designed; these expressions do not belong to the sphere of discourse of ‘piece of wood’.

Furthermore, associated with a sortal term, there is typically a series of persistence conditions which tell when an instance of the sortal starts or ceases to satisfy the sortal or instantiate the kind. A statue would typically cease to be a statue if it were hammered flat; a paper aeroplane would typically start to be a paper aeroplane if a piece of paper is folded in a particular manner; and a kitten would typically stop being a kitten when it turns a certain age.

Based on the individuation conception of sortals, it seems hard to deny that an object typically satisfies more than one sortal term. Let me firstly elaborate by giving some examples: a cat is not only a cat but (possibly) also a kitten as well as a living creature; a wooden table is not only a table but also a yellow table as well as a piece of wood; a truck is not only a truck but also a car, a machine, an integrate of parts, a large collective of metals, as well as a summation of atoms. The reason for this proposition is that all these sortals, on the one hand, are good candidates for the answer to the ‘What is it?’ question, and, on the other hand, can be meaningfully attributed to the objects at issue.

With these features of sortals at hand, we can open the discussion of the puzzle of coincidence which, I think, can be accounted for by these features of sortal terms.

3. The puzzle of Coincidence

Suppose I have a full drawer with all sorts of items including a box of markers. Without noticing this fact, I buy a new box of markers and find there is no room in the drawer. In this circumstance, it would be completely nuts to talk myself into placing the box of markers in the drawer (more precisely, in the region where the former box lies), by telling me ‘Distinct objects can occupy

the same region at the same time. So no worries just put it there!’ After all, metaphysical principles and laws of physics stipulate that objects are extended and extended entities cannot pass through each other.

A series of cases has been proposed to demonstrate the view that distinct objects can occupy the same place at the same time. These cases principally take two ways of argumentation:

- (1) Suppose on an otherwise empty table there is a piece of paper. Then a kid comes to the table and folds this piece of paper into a paper aeroplane. Let us call the original piece of paper ‘Piece 1’ and the paper plane ‘Aeroplane’. In the region Aeroplane occupies, it seems plausible that there is also a piece of paper. Let us call it ‘Piece 2’. Now, because a piece of paper can survive mere changing in shape, without any quantity of matter being lost, it appears plausible that Piece 2 is Piece 1; and because Aeroplane only comes into existence when the kid finished her folding while Piece 1 exists before that, it seems evident that Piece 1 is not Aeroplane. Given the premises above, we are pushed to the conclusion that Piece 2 is not Aeroplane.

The main argument can be expressed in the following form:

Premise 1: Piece 2 = Piece 1

Premise 2: Piece 1 \neq Aeroplane

Conclusion: Piece 2 \neq Aeroplane

In this form of argument for the non-identity of the apparently coinciding objects, different histories of the objects at issue are invoked, and several identity relations are employed to come to the bizarre conclusion.

- (2) The second form of argument does not specifically invoke temporal properties. Consider the philosopher’s favourite toy example: statue/ piece of copper. On an otherwise empty table, there stands a copper statue. It seems true to say that at the place where the statue stands, there is also a piece of copper. Now there is the following argument:

Premise 1: The statue is Romanesque / well-made / valuable / insured / admired.

Premise 2: The piece of copper is not Romanesque / well-made / valuable / insured / admired.

Conclusion: The statue is not the piece of copper.

A more general form of this argument has been given by King (2006), which he calls The Master Argument (MA):

$\Phi(a)$

$-\Phi(b)$

Therefore, $(a \neq b)$,

in which the predicate ' Φ ' expresses a property and the variables ' a ' and ' b ' are both singular terms.

The argument goes through by the principle of the indiscernibility of identicals (also called Leibniz's Law), which says if a is identical to b , then any property possessed by b is possessed by a . In the statue case, since the statue's property being Romanesque, is not one of the properties of the piece of copper, it follows that the piece of copper is not identical with the statue. This conclusion implies that different objects can occupy the same place at the same time.

In either form of argumentation, we seem to be compelled to accept the position that distinct objects can coincide at the same time. Given that the conclusion of these lines of argument clearly contradicts the view that a space cannot be shared by different objects, we meet the metaphysical puzzle of coincidence.

As I see, in many cases, the reason why the first premise is true while the negated second premise is false is that the predicates in the first premise can be meaningfully attributed to the sortal term, i. e. the predicate belongs to the sphere of discourse of sortals, yet the predicates in the second premise typically cannot be meaningfully attributed to the sortal terms, because they do not belong to the sphere of discourse of the sortals in premise 2.

There seem to be cases of the second form of argument for non-identity that can be readily explained away by considering the characteristics of sortals, and our intuition against place-sharing could thereby be rescued.

Consider the following argument:

The cat survives when it turns 10.
 The kitten does not survive when it turns 10.
 Therefore, the cat is not the kitten.

Clearly, something goes wrong in the second premise. The kitten does survive when it turns 10. This is true because the sortal 'kitten' is a phase sortal, that is, a sortal denoting a kind that an object can acquire and lose without beginning or ceasing to exist. The solution to this easy case would also relieve us, at least to a little extent, from the pluralist position that objects typically coincide.

Besides, if we mean by 'existence' that an object satisfies a certain description (either a property or a kind), then it is very plausible to say that "‘there is a kitten on the table’ is true". If we mean by 'existence' that there is such a thing in nature, independent of our conceptual scheme, then we need to address the question of whether 'kitten' is a natural kind, whose answer seems to be false.

But there are other issues about sortals that interest me more.

Two related issues concern the puzzle of coincidence: most of the cases of coincidence would parish if there are no ordinary objects - if eliminativism is true, the statements in the arguments would, at least in a strong reading, be false, such that the whole argument is simply unsound. We would have an even more wild world view if there were not only ordinary objects but also extraordinary objects such as snowdiscalls (an example from Sosa (1987), the piece of snow having a shape between roundness and disc-shape) as well as the summation of the washing machine in the garage and the cow in the field – if permissivism is true, the arguments by Leibniz's Law would bring conclusion such as at the place where the statue lies, there are really infinitely many objects there at the same time, that strikes most of us as unacceptable.

But whether ordinary as well as extraordinary objects exist, I will argue, turns on our understanding of the relation among sortals, kinds and objects. We claim there is such-and-such a thing, very importantly because there is a

corresponding sortal term within our grasp. The sortal terms are something available to us. Even though it does not follow that there are for each sortal an entity as kind, it seems to imply that there is something in nature that allows us to have these sortal terms. And if, as Lowe (2009) argues, kinds and objects are interdependent and not reducible to each other, the natural kinds seem to be able to give us an answer to what really exists.

To summarize, in this section, I have tried to show that the puzzle of coincidence is not a metaphysical problem in the sense in which it pertains to the underlying nature of the world. More specifically, if the arguments for non-identity go through, I suggest that it does not necessarily mean that *reality* allows objects to occupy the same region at a certain time; and if our intuition about place-sharing persists, I hold that it does not necessarily mean there are *really* such objects in the universe. Answers to these questions have to wait until a more comprehensive understanding of sortals is at hand.

References

- Almotahari, Mahrad (2014) "Metalinguistic negation and metaphysical affirmation", *Philosophical studies*, 167, 497-517.
- Burke, Michael B. (1992) "Copper statues and pieces of copper: a challenge to the standard account", *Analysis (Oxford)*, 52, 12-17.
- Burke, Michael B. (1994) "Preserving the Principle of One Object to a Place: A Novel Account of the Relations Among Objects, Sorts, Sortals, and Persistence Conditions", *Philosophy and phenomenological research*, 54, 591-624.
- Fine, Kit (2003) "The Non-Identity of a Material Thing and Its Matter", *Mind*, 112, 195-234.
- King, Jeffrey C. (2006) "Semantics for Monists", *Mind*, 115, 1023-58.
- Locke, John. (2008[1690]) *An essay concerning human understanding. Abridged with an introduction and notes by Pauline Phemister* (New York : Oxford University Press: New York).
- Lowe, E. J. (2006) *The four-category ontology a metaphysical foundation for natural science*. Oxford: Clarendon: Oxford.
- Lowe, E. J. (2009) *More kinds of being: a further study of individuation, identity, and the logic of sortal terms*. Chichester, U.K. Malden.
- Mooney, Justin "A Phasalist Approach to Coincidence Puzzles", *The Philosophical quarterly*.
- Mooney, Justin (2023) "Criteria of identity without sortals", *Noûs*, 57, 722-39.

Mooney, Justin (2024) "The Matter of Coincidence", *Pacific Philosophical Quarterly*, 105, 98-114.

Sattig, Thomas (2015) *The double lives of objects: an essay in the metaphysics of the ordinary world*. Oxford University Press: Oxford.

Sosa, Ernest (1987) "Subjects Among Other Things", *Philosophical perspectives*, 1, 155-87.

Wiggins, David (1980) *Sameness and substance*. B. Blackwell: Oxford.

The Modal Argument for the Anchoring Model in Social Metaphysics

Samuele Chilovi (Madrid, Spain)

Abstract

Extended Abstract. Recent years have witnessed a surge of interest in the nature and structure of social reality. As part of this development, attempts have been made to elaborate general models of the explanation of social entities, and specifically of social facts, such as facts about the instantiation of social properties and relations by particular items. Particular social facts are plausibly viewed not as fundamental constituents of reality, but rather as derivative and dependent upon more basic entities. This raises the question of how they are metaphysically explained: what explains them, and what is the nature of the explanatory relation(s) that ties them to their constitutive determinants.

Brian Epstein (2015, *The Ant Trap*, Oxford: OUP) has distinguished two competing models of the explanation of social facts. On one account (defended, e.g., in A. M. Griffith, 2017, 'Social Construction and Grounding', *Philosophy and Phenomenological Research* 97: 393-409; J. Schaffer, 2019, 'Anchoring as Grounding', *Philosophy and Phenomenological Research* 99: 749-67), social facts are partly grounded by facts about the particular in question, and partly by social rules, which, in turn, are themselves grounded by their determinants (e.g., social practices, collective actions and intentions). If, say, Esther has the right to vote in the US elections, this fact is partly grounded in other facts about Esther – that they're an 18-year-old US citizen, say – and is also partly grounded in the existence of a social (legal) rule – that 18-year-old US citizens have the right to vote in the elections – itself grounded in some underlying practices, acts or mental states. This account has been variably called "conjunctivism" and "grounding-only model" (Epstein 2015; 2019a, 'Anchoring versus Grounding: Reply to Schaffer', *Philosophy and Phenomenological Research* 99: 768-81; 2019b, 'Replies to Hawley, Mikkola, and Hindriks', *Inquiry* 62: 230-46), to emphasize that social facts are grounded in two sorts of ingredients (rules as well as facts), and that only grounding relations are needed to model the explanation of social facts.

In contrast with this model, Epstein has developed an alternative account that, in addition to grounding relations, countenances an additional relation of metaphysical determination he calls "anchoring". On Epstein's anchoring-grounding model, social practices (and the like) set up, by anchoring, general conditions for how social kinds are grounded. Then, facts about the satisfaction of these conditions fully ground social facts. Therefore, for example, law-making practices anchor the conditions for having the right to vote in the US elections, whereas facts about people's satisfaction of these conditions fully ground their having the right to vote. Here, social (or legal) rules play no role in the explanation of social facts. Rather, they are mere summaries of how social properties are grounded across modal space. And a relation of metaphysical determination distinct from grounding is taken to be necessary to model the explanatory structure of social reality.

Epstein (2015, 2019a, 2019b) has raised an influential and powerful argument against the grounding-only model and in favor of his own alternative account. Social kinds, he submits, can be modally flexible in ways that grounding cannot accommodate, in that they can be instantiated at worlds where the practices that set them up are absent. For instance, the kind war criminal was anchored in social practices that took place in Europe in the XX century,

with the enactment of the Geneva Conventions in 1949. Yet we can meaningfully ask whether Genghis Khan, having committed atrocities that count as war crimes by the Conventions, would still be a war criminal at a world that ended in 1500, where the practices that set up the Conventions are absent. Epstein submits that he would. This social kind is universal in that, once set up at the actual world, establishes conditions that can be met even at worlds where no anchors are present. If Genghis Khan satisfies these conditions at a world that lacks them, he still counts as a war criminal there. Grounding, however, is a world-bound relation: it only relates entities that exist or obtain at the same possible world. Therefore, it cannot hold between anchors and social facts when these are modally disjoint. So, we need a distinct relation to capture the link between these two.

One possible reply is to deny that social kinds can allow for this sort of modal freedom, so that social facts cannot be modally disjoint from their anchors. Another possible reply is to concede that the phenomenon is real – that there are cases of “modal exportation”, as Schaffer calls them (see Schaffer 2019) – yet maintain that this does not warrant the introduction of a *sui generis* determination relation distinct from grounding, for grounding can, in certain cases, hold between entities existing at different possible worlds.

Both of these solutions have costs. On the former, it is worth noticing that there need only be one case of a kind that is modally flexible in the way that Epstein envisions for the problem to emerge, so taking issue with Epstein’s specific example doesn’t preclude that other kinds might to the trick. The latter reply, on the other hand, adopts the claim that grounding can be a cross-world relation. This not only is dubious in itself, but also raises the question of when and why this should be possible.

A more appealing solution is one that avoids incurring these costs. Here, I argue that such a solution can be found. The modal argument says that grounding and anchoring are distinct relations because they have different modal profiles: anchoring, but not grounding, can hold between different possible worlds. That anchoring holds between different worlds is meant to be supported by the existence of social facts that are modally disjoint from their anchors. The fact that Genghis Khan is a war criminal, for instance, can obtain at worlds where no enactment of the Geneva Conventions (and the like) takes place. The problem, however, is that the anchoring-grounding framework does not countenance the existence of a cross-world relation either, not even in cases such as these. In the Khan case, for instance, what happens is that at the actual world social practices about enactments anchor the social kind war criminal, whereas at some other possible world Khan’s committing the atrocities grounds his being a war criminal. So what we are left with is not a cross-world relation, but rather two intra-world relations holding at different worlds.

Translating the Tractatus and Tractarian Ethics

Jeffrey Patrick Colgan (New Orleans/ Iowa City, USA)

Abstract

As the number of *Tractatus* translations have increased, translators have sought to justify and distinguish their translations by enumerating the specific considerations that have guided their work. This paper treats the consideration of opacity, which has been notably absent from the explicit considerations of *Tractatus* translators. Opacity, as I use it, is a characteristic of texts or passages from texts that demand substantial interpretive work, allow for repeated interpretations, and resist being “settled.” Opaque passages are hermeneutically unstable. This paper argues that opacity is an unavoidable and crucial consideration when translating the *Tractatus*. As opacity in the *Tractatus* is not uniform, I first identify three types of Tractarian opacity: defeatist, esoteric, and systematic. To substantiate these types of Tractarian opacity, I present historical evidence from Wittgenstein’s correspondence and consider various candidate opaque passages. Each of these three types of opacity present particular challenges to any translator of the *Tractatus*, and I engage with opacity as a specifically translational problem, situating it within the history of translation studies and arguing that it has been a perennial concern for translators in certain traditions. How one deals with opacity in the *Tractatus*, I suggest, is ultimately dependent upon how one interprets the ethical thrust of the text. I conclude the paper by suggesting that translating the *Tractatus* (and not just the explicitly ethical passages) requires one to take a stand on Tractarian ethics.

1. Making a Case for the Consideration of Tractarian Opacity

With the expiration of the copyright on Wittgenstein’s *Tractatus Logico-Philosophicus*, a flurry of English translations have been published or are currently in the works. Recent translations include those from Michael Beaney (2023), Alexander Booth (2023), Damion Searls (2024), and a forthcoming edition from David Stern, Joachim Schulte, and Katia Saporiti. These join the well-known translations by Ogden and Ramsay (1922) and Pears and McGuinness (1961), as well as other translations that have often accompanied commentaries on the text, such as those from Daniel Kolak (1998) and Duncan Richter (2021). As the number of *Tractatus* translations have increased, translators have sought to justify and distinguish their translations by enumerating the specific considerations that have guided their work. Booth’s 2023 translation seeks to foreground the text’s poetic and literary qualities, presenting what Jan Zwicky in the introduction to that edition calls its “musicality” and “lyric nature” (Booth, p. xx). Searls’ 2024 translation strives for “normalcy” and colloquial language usage (Searls, p. xxxvi, xxxviii). Beaney, in explaining the approach for his commendable 2023 translation, goes even further and offers seven guiding considerations (Beaney Draft, pp.

2-3). Notably absent from these translators' explicit considerations—though undoubtedly an aspect that every translator grapples with—is the opacity that characterizes many of the *Tractatus*' passages, its gnomic quality. Opacity, as I use it, is akin to Friedrich Schleiermacher and Lawrence Venuti's use of the term foreignness, and about which I'll have much to say below; it is a characteristic of texts or passages from texts that demand substantial interpretive work, allow for repeated interpretations, and resist being "settled." Opaque passages are hermeneutically unstable. Opacity is distinguished from ambiguity in that ambiguity allows for multiple successful interpretations, whereas opacity leaves open the option of there being no successful interpretation. An opaque passage may be ambiguous, but it may also simply be enigmatic.

This paper seeks to take the consideration of opacity and clean it up, give it a haircut and a shave, so that it is not something rejected out of hand nor simply gestured to, but considered seriously. Opacity in the *Tractatus* is not uniform, though, and in Part II of this paper I offer a taxonomy of sorts, identifying three types of Tractarian opacity: defeatist, esoteric, and systematic. To substantiate these types of Tractarian opacity, I present historical evidence from Wittgenstein's correspondence and consider various candidate opaque passages. Each of these three types of opacity present particular challenges to any translator of the *Tractatus*. In Part III, I engage with opacity as a specifically translational problem, situating it within the history of translation studies and arguing that it has been a perennial concern for translators in certain traditions. How one deals with opacity in the *Tractatus* is ultimately dependent upon how one interprets the ethical thrust of the text. Thus in Part IV, I connect the consideration of opacity with the debates about how to interpret the *Tractatus* as an ethical text. My concluding suggestion is that translating the *Tractatus* (and not just the explicitly ethical passages) requires one to take a stand on the text's ethical thrust; the ethical is present throughout the text, wherever there is opacity.

2. Opacity in the *Tractatus*

I take it as uncontroversial that there are many passages in the *Tractatus* that can be characterized as opaque. Labelling a passage opaque does not imply

that it must resist interpretation; in fact, many readers of the *Tractatus* are perhaps quite confident in their ability to offer an interpretation for any and all passages of the text. What opacity means in this context is rather the demand for substantial interpretation on behalf of the reader and translator and the latitude of plausible interpretations offered. The number of readings of the *Tractatus* that have been produced and the busy realm that is *Tractatus* interpretation both suggest the presence of productive opacity throughout the text. I offer three types of opacity that can be found in the *Tractatus*: defeatist, esoteric, and systematic. In what follows, I treat each in turn and offer examples when possible.

A passage is characterized by defeatist opacity when the author has struggled to make clear her thoughts. That Wittgenstein struggled to adequately—especially to his standards—express his thoughts is well-documented. In a 1922 letter to Ogden, Wittgenstein writes:

When I had finished the book *roughly* there remained certain prop[osition]s – about a hundred – about which I was doubtful whether I should take them in or not. These prop[osition]s were – partly – different versions of those now contained in the book; for it had often happened that I had written down a prop[osition] in many different forms, when the same thought had occurred to me in different ways during the long time I worked at that business. (LO, p. 46)

Wittgenstein goes on to call these inadequate attempts to express his thoughts “less clear than the rest of my prop[osition]s” and compares them to “the shavings and sawdust and other rub[b]ish” left over from the construction of a table (LO, p. 46). The same sentiment is present in the preface to the *Tractatus*, when Wittgenstein acknowledges that thoughts can be expressed better or worse and admits his awareness (*bewusst*) of falling short of what is possible because of a weakness in his ability (*TLP*, Preface).

The correspondence between Wittgenstein and Ogden over the translation of the *Tractatus* contains numerous instances where Wittgenstein intimates the difficulty he experienced in attempting to express his thoughts in the original German and his willingness to deviate from the German in the English translation, if such deviations could achieve greater clarity of his initial

thoughts. Even at this stage, when the *Tractatus* is being translated into English, Wittgenstein still appears to be at times struggling to express himself. In Wittgenstein's notes to Ogden on *TLP* 6.4312, he attempts to explain to Ogden what the original German sought to convey by means of a lengthy example (LO, p. 35). For 3.328, Wittgenstein deviates from the German original to make the point more explicit (LO, p. 25). 4.0141 is added into the *Tractatus* from a collection of supplementary remarks (the shavings and sawdust) as a supplement to 4.014, presumably to better explain the work that 4.014 is trying to do (LO, p. 26). These sections of the *Tractatus* serve not only to illustrate the difficulty that Wittgenstein faced in expressing his thoughts but also point out specific passages that, at least to Wittgenstein, retained a level of opacity.

The second type of opacity in the *Tractatus* is esoteric opacity. Esoteric opacity is the opacity that characterizes passages with both a surface and deeper meaning or an intended meaning for a specific group of readers that will be unacknowledged by readers outside that group. Frank Ramsay in a 1923 letter to his mother echoes the *Tractatus*' preface and reports that

His idea of his book is not that anyone by reading it will understand his ideas, but that some day someone will think them out again for himself, and will derive great pleasure from finding in this book their exact expression. (LO, p. 78).

In the same letter, Ramsay relays that “some of [Wittgenstein's] sentences are intentionally ambiguous having an ordinary meaning and a more difficult meaning which he also believes” (LO, p. 78). Russell, after calling Wittgenstein a “complete mystic” (Beaney 2023, p. lxiii), reports something similar in a 1912 letter:

I told him he ought not simply to *state* what he thinks true, but to give arguments for it, but he said arguments spoil its beauty, and that he would feel as if he was dirtying a flower with muddy hands.” (Beaney 2023, p. lxxxvi)

Wittgenstein is deliberately holding back in his remarks, unwilling to make his point clearer, or purposefully offering ambiguities.

Related to esoteric opacity is a pedagogical aspect of Wittgenstein's writings. As the *Tractatus*' preface states, it is not a textbook; rather, as Beaney suggests, it is "an exercise book: a text that has to be thought into a book for oneself" (Beaney 2023, p. lxxxvii). Unlike the esoteric writings of those seeking to avoid arousing state or religious censors, Wittgenstein's esotericism might best be understood as due to his intent to avoid proffering dogmas and to make his readers think—as well as due to aesthetic and stylistic concerns.

Offering clearcut examples of *Tractatus* passages with esoteric opacity is intrinsically problematic. Any esoteric passage—because it is esoteric—will be grasped by some and not by others. However, Wittgenstein in the *Tractatus* and in his discussions and correspondence illustrates a willingness to conceal some of his intentions in the *Tractatus*.

A third type of opacity in the *Tractatus* is systematic opacity. This is the opacity characteristic of passages that offend the account of meaning offered in the *Tractatus*; this is the opacity of nonsense. Whether understood as strings of names that fail to abide by the combinatorial possibilities of the objects to which they refer or as sentences without sense and significance that do not permit any scrutiny of their constituents, and further whether this nonsense is understood as mere nonsense or substantive nonsense, this opacity is that which results when language use attempts to go beyond its limits (*TLP*, 4.02-4.03; *TLP*, 3.3, 3.4). Present in the preface of the *Tractatus* as well as the 6.4s and 6.5s, this is what Wittgenstein calls in his letter to Ludwig von Ficker "gassing" or in the "Lecture on Ethics" as "essential nonsensicality" and language that leaves behind factual expression and attempts to "go beyond the world" (von Wright, p. 83; *LE*, pp. 50-51). It is systematic opacity because it is not contingent upon having "not yet found the correct expressions" but because these instances attempt to transgress the limits of language (*LE*, p. 50). Though this is the deepest sort of opacity found in the *Tractatus* because it is, according to the *Tractatus*, unavoidable, this is the opacity with which most readers are familiar.

Representative passages in the *Tractatus* include 6.42 and 6.432, with their discussion of the higher (*Höheres*); the discussion of ethics in 6.421 and 6.422; 6.43, with its discussion of good or bad willing; the discussion of death as a personal experience in 6.431 and 6.4311; the mention of God in 6.432; *that* the

world is (*dass sie ist*) in 6.44; the feeling of the world as a limited whole in 6.45; the solution to the problem of life in 6.521; and the ineffable (*Unaussprechliches*) and the mystical (*das Mystische*) in 6.522.

3. Opacity and Translating the *Tractatus*

Opacity presents specific challenges to the practice of translation generally and the translation of the *Tractatus* specifically. Opaque passages show any naive insistence on equivalence to be wrongheaded, which of course conflicts with the explicit treatment of translation in the *Tractatus* (TLP, 3.343, 4.025). Without equivalence, the translator's craft reaches its height, and it is here at these sites of opacity that any specific target text demonstrates its quality and a given translator proves her mettle. To navigate these challenges in a principled way, translators often offer guiding considerations for their translations. In offering such a set of considerations, translators are forced to reflect upon difficult questions about who will be the audience, to what extent rendering the target text accessible to that audience limits the originality and idiosyncrasies of the source text, and to what extent the source text is, in pursuit of fluency, distorted according to a particular interpretation. In what follows I consider the guiding considerations offered by some of the more recent *Tractatus* translators and argue that opacity itself should be among those considerations that guide the translation of a text like the *Tractatus*.

As mentioned above, Beaney identifies seven considerations that guided his translation. He glosses them as follows (Beaney Draft, p. 2-3):

1. Faithfulness: The text is all that is translated.
2. Correspondence: What is translated are sets of related expressions.
3. Consistency: An ideal translation of a German expression is a single English expression.
4. Multiplicity: A translation is a reflection of the full sense or senses of an expression.
5. Flexibility: A translated proposition is a transformed function of the translation of its parts.
6. Fluency: The general form of a translation is: this is what the author would have said in the language of the translator.

7. Invisibility: Where invisibility [of the translator] is impossible, speaking is imperative.

Stern broadly agrees with Beaney's stated considerations; however, in his own translation of the *Tractatus* he differs from Beaney by putting less weight on the principles of correspondence and consistency (Stern Draft, p. 7).

Not only is opacity absent from both Beaney and Stern's stated considerations, it seems to oppose certain considerations—most obviously fluency and the invisibility of the translator. Fluency especially can lead to a neglect of opacity in the source text, running the risk of distorting the original and eradicating valuable idiosyncrasies.

Opacity as a consideration is, I take it, present in the translation practices of both Beaney and Stern; it is just not made explicit. Further, opacity is not a new or remote concern for translators and scholars of translation studies. Friedrich Schleiermacher in his 1813 lecture "On the Different Methods of Translating" presents an opposition between the foreign (source text) and the domestic (socio-linguistic group that the target text will address). For Schleiermacher "a foreign language will always be available in fragmentary form only: the exegete has not grown up in and with that idiom and can therefore only ever grasp it partially and imperfectly, as an outsider" (Hermans, p. 27). The best a translator can do is to convey "to readers unfamiliar with the foreign language that particular sense of the foreign as it inhabits this specific work by this individual writer and as the translator...has apprehended it. [Thus] foreignness...enters the translating language" as an inescapable aspect of the source text that evades full intelligibility (Hermans, p. 29). Opacity is insurmountable in translation.

Translator and translation studies scholar Lawrence Venuti picks up this line of thought and criticizes (primarily Anglo-American) philosophers and translators for adhering to a "preference for fluency, immediate intelligibility, the illusion of transparent communication" (Venuti, p. 116). Domestication is, of course, unavoidable and in fact "necessary if the foreign text is to become intelligible and interesting to domestic readers" (Venuti, p. 114). However, for Venuti the dominant approach results in translations that unavoidably "revert to a domestic standard...stylistic canon or...interpretation" (Venuti, p. 106). In

its stead, he offers the strategy of “reading for the remainder.” He explains his strategy as one that “continues to be philosophical, engaged in conceptual analysis, but now made more literary, concerned with the formal properties of language,” and one that “requires a reformulation of the notion of accuracy, a broadening that takes into account both the foreign text and domestic readers” (Venuti, p. 114-115).

Opacity as I use the term relates to this notion of foreignness in Schleiermacher and Venuti, though without the emphasis on large-scale socio-linguistic communities or the subtext of national character. Rather, the emphasis is on the idiosyncrasies of specific authors and texts, especially the unconventional uses of language and textual form that demand a departure from our everyday and expected uses of language. In the *Tractatus* specifically, it is a consideration to aid in the avoidance of flattening the text by subsuming it wholesale into a philosophical tradition that it is, in part, criticizing. The history of the translation of the *Tractatus* into English is, in many ways, a history of confronting its opacity.

4. Opacity and Tractarian Ethics

It is uncontroversial that the translations of the explicitly ethical passages are often influenced by the translators’ interpretations of Tractarian ethics. A most egregious example of this comes from Kolak’s 1998 translation, where “*transcendental*” in 6.421 is rendered as “transcendent.” But this latitude of translation can be found in many passages. Consider the various translations of “*zeigt*,” “*Unaussprechliches*,” and “*allerdings*” in 6.522:

Deutsch

Es gibt allerdings Unaussprechliches. Dies *zeigt* sich, es ist das Mystische.

Ogden/Ramsay

There is indeed the inexpressible. This *shows* itself; it is the mystical.

Pears/McGuinness

There are, indeed, things that cannot be put into words. They *make themselves manifest*. They are what is mystical.

Kolak

The inexpressible indeed exists. This *shows* itself. It is the mystical.

Beaney

There is, though, the ineffable. This *shows* itself, it is the mystical.

Searls

Of course there are things that cannot be spoken. They *show* themselves; they are mystical.

Booth

There is indeed the ineffable. It *shows* itself; it is the mystical.

Stern

Some things can't be said, though. They *show* themselves; they are what is mystical.

One can find similarly stark differences in the preface, 6.42, and 7.

However, I want to conclude by suggesting that one's interpretation of Tractarian ethics—the so-called ethical point of the book—does not only bear upon the explicitly ethical passages but on the opaque passages throughout the text. The thought goes as follows: the three types of opacity treated in this paper can be located throughout the *Tractatus*—not only in the preface, 6.4s, 6.5s, and 7. Further, any specific instance of opacity will in most cases be underdetermined with respect to which kind of opacity it is. As such, each instance of opacity is ethically-relevant in that it is either itself nonsensical or entangled with those opaque passages that are nonsensical. How a translator (or even a reader) interprets such instances of opacity depends upon how she interprets the ethical thrust of the *Tractatus*—e.g., resolute, ineffabilist, or like Yaniv Iczkovitz as a spiritual exercise of purification (see Iczkovitz 2012; for general overviews see Christensen 2024 and Bronzo 2012). Thus, the ethical thrust of the text permeates the whole of any *Tractatus* interpretive project, and this point holds irrespective of one's particular interpretation of Tractarian ethics. Translating (and interpreting generally) the opaque passages throughout the *Tractatus* requires one to take a stand on Tractarian ethics.

References

Beaney, Michael (Draft) "Translating the *Tractatus*," unpublished draft of an in-progress paper.

Bronzo, Silver (2012) "The Resolute Reading and Its Critics: An Introduction to the Literature," *Wittgenstein-Studien*, Vol. 3, 45-80.

Christensen, Anne-Marie Søndergaard (2024) *Wittgenstein and Ethics*, Cambridge: Cambridge University Press.

Hermans, Theo (2019) "Schleiermacher," in: *The Routledge Handbook of Translation and Philosophy*, New York: Routledge, 17-33.

Iczkovitz, Yaniv (2012) *Wittgenstein's Ethical Thought*, Basingstoke: Palgrave Macmillan.

Stern, David (Draft) "On Translating the *Tractatus* Consistently," unpublished draft of an in-progress paper.

Venuti, Lawrence (1998) *The Scandals of Translation: Toward an Ethics of Difference*, New York: Routledge.

von Wright, Georg (1982) "The Origin of the *Tractatus*," in: *Wittgenstein*, Oxford: Oxford University Press, 63-109.

Wittgenstein, Ludwig (2014) *Lecture on Ethics*, ed. with commentary by Edoardo Zamuner, Ermelinda Valentina Di Lascio, and D.K. Levy, Chichester: Wiley.

Wittgenstein, Ludwig (1973) *Letters to C.K. Ogden with Comments on the English Translation of the Tractatus Logico-Philosophicus*, ed. with an introduction by G.H. von Wright, Oxford: Blackwell.

Wittgenstein, Ludwig & Anscombe, G.E.M. & Hacker, P.M.S. & Schulte, J. (2009) *Philosophical Investigations*, Chichester, Wiley.

Wittgenstein, Ludwig & Beaney, Michael (2023) *Tractatus Logico-Philosophicus*, Oxford: Oxford University Press.

Wittgenstein, Ludwig & Booth, Alexander (2023) *Tractatus Logico-Philosophicus: The New Translation*, London: Penguin.

Wittgenstein, Ludwig & Kolak, Daniel (1998) *Wittgenstein's Tractatus*, Mountain View: Mayfield.

Wittgenstein, Ludwig & Ogden, C.K.; Ramsay, Frank (1922, 1999) *Tractatus Logico-Philosophicus*, Mineola: Dover.

Wittgenstein, Ludwig & Pears, D.F.; McGuinness, B.F. (1961) *Tractatus Logico-Philosophicus*, New York: Routledge.

Wittgenstein, Ludwig & Richter, Duncan (2021) *Wittgenstein's Tractatus: A Student's Edition*, Lanham: Lexington.

Wittgenstein, Ludwig & Searls, Damion (2024) *Tractatus Logico-Philosophicus*, New York: W.W. Norton & Company.

Wittgenstein, Ludwig & Stern, David; Schulte, Joachim; Saporiti, Katia (Forthcoming) *Tractatus Logico-Philosophicus*, In-progress translation.

Acknowledgement

I would like to thank David Stern, Joachim Schulte, and Katia Saporiti for permission in using their unpublished and in-progress translation of the *Tractatus*. I would like to thank David Stern specifically for his guidance in pursuing this project.

Vaidya's Method of Variation in Imagination Revisited

Violeta Conde, Concha Martínez-Vidal (Santiago de Compostela, Spain)

Abstract

In their famous paper *The epistemology of modality and the problem of modal epistemic friction* (2018), Vaidya & Wallner propose to switch the debate in the epistemology of modality from possibility-first vs. necessity-first approaches to possibility-first vs. essences-first approaches. This decision obeys the identification of the problem of *modal epistemic friction* for some accounts of epistemology, according to which are essentialist theses which underpin those accounts. If that thesis is correct, the epistemology of modality should be a special case of the epistemology of essences. In this vein, Vaidya proposes a way to understand essences—and, therefore, to gain modal knowledge—which is based in what he calls *variation in imagination*: the essential properties of an object would be those that remain invariant through a method of varying properties in imagination from time to time. However, since this method is closed under logical consequence, the notion of essence involved is, in Finean terms, consequentialist and not constitutive. Therefore, our aim in this talk will be to modify Vaidya's proposal suggesting a way to capture constitutive essences starting from consequentialist essences using the notions of *grounding* and *relevance*.

In their famous paper *The epistemology of modality and the problem of modal epistemic friction* (2018), Vaidya & Wallner propose to switch the debate in the epistemology of modality from possibility-first vs. necessity-first approaches to possibility-first vs. essences-first approaches. This decision obeys the identification of the problem of modal epistemic friction for some accounts of epistemology of modality such a Williamson's counterfactual theory (2007), Yablo (1993) and Chalmers (2002) conceivability theories or Kripke (1971) deduction theory. Roughly put, this problem consists in that the former accounts rely on essentialist propositions. If Vaidya & Wallner thesis is true, the epistemology of modality would be just a special case of epistemology of essences. We will assume the former statement for the purposes of this paper.

But how can we know essences? It seems that to know something there should be a connection between the subject and the truth-maker (knowledge will be, therefore, incompatible with epistemic luck) and it does not seem we have that connection in case of essences. This could be viewed as the modal version of Benacerraf's problem in epistemology of mathematics. However, in a paper entitled *Understanding and essence* (2010) Vaidya makes a gambit to solve that dilemma and argues that we should address the debate possibilist-first vs essences first making a difference between “knowing essences” and “understanding essences”. Even if both verbs—“know” and “understand”—are

viewed as factive, the first one is incompatible with epistemic luck, whereas the second is not. What Vaidya calls “objectual understanding” allows him to accept that we can accidentally arrive at the truth without having any causal connection between the subject and the truth-maker, and, therefore, to avoid what he calls the Meno paradox when using his method of variation in imagination to understand essences. The basic idea in Vaidya’s proposal is that “we can make a judgement about the essential properties of an object by varying properties of the objects in imagination, and seeing which vary and which do not” (Vaidya, 2010: 820). The variant properties would be the accidental ones, whereas the invariant would be the essential ones. However, the method of variation in imagination is closed under logical consequence, so it produces essential properties that are not constitutive, but consequential, to use Finean (2012) terms. Fine argues that is possible to arrive to properties that are essential in the constitutive way starting from properties that are essential in the consequentialist way and filtering them out. Our aim in this paper is to pursue this Finean strategy to improve Vaidya’s method of variation in imagination by suggesting that the needed depuration can be done applying the notions of grounding (Fine, 2012) and relevance (Schnieder 2018; Hirèche 2023).

The method of *variation in imagination* proposed by Vaidya is based in Husserl’s method of *eidetic variation*. He argues that through it we can *understand* what essences are and consider them as an entry point to our knowledge of modality. To illustrate this method, let’s think of Aristotle in first instance: in a first moment, namely, t_1 , we can think of him having certain properties, such as “being human”, “being a philosopher”, “having two legs”, “being born in Stagira” etc., in a second moment, t_2 we can think of him as “being human”, “being a musician”, “having one leg”, “being born in Athens”, etc.; in a third moment, t_3 , we can think of him as “being human”, “being a soldier”, “having no legs”, “being born in Elea”, etc. In any case, we cannot think of him as not being human, so it will be in the essence of Aristotle the fact of being human (and every property which remains invariant through the process of variation in imagination). In some way, this is similar to Kripke’s view in *Naming and Necessity* (1980) when he says that a property is essential to an object if and only if we cannot imagine a world in which that object does not have that property.

Vaidya (2010) is aware that this method leads to a version of Meno paradox if we think that we are doing is *knowing* essences:

Intuitively, the problem is that one cannot preserve an object across a series of transitions unless they know which properties they can alter. But knowing which properties one can alter would require that one already, either implicitly or explicitly, knows which properties are essential and which are not (Vaidya 2010: 822).

Vaidya solves this worry adducing that we are really doing is *understanding* essences and not *knowing* them, so our understanding is compatible with epistemic luck. But our target here is other. This procedure is closed under logical consequence; thus, i.e, for every disjunction one of whose members is “Aristotle is human”, for example, could be considered as forming part of the essence of Aristotle. The problem is not that we cannot capture every property in the essence of Aristotle, for Vaidya acknowledges that this procedure cannot be complete:

Second, when one considers a set of scenarios, the set of scenarios may not be complete. Note in the example above, the initial set of properties considered does not constitute a complete list of properties of O. However, it is not necessary that in coming to have a justified belief, knowledge, or understanding of O's essence that one examines every property of O. What is required is that the set of properties be adequate and representative in an appropriate sense, not that the set be complete of all properties of the object. Completeness or exhaustiveness as a requirement on the set of properties considered would be over demanding, and render any finite exercise of variation epistemically irrelevant (Vaidya 2010: 821).

We agree with Vaidya that the procedure, in order to be epistemically relevant, should not be complete. Still, the problem is that the set of properties cannot be “adequate and representative” if the procedure is closed under logical consequence. We want to avoid conclusions such as “being human or the moon is made up of cheese” or “not being Plato” being in the essence of Aristotle. If we go back to the example of Aristotle, we can appreciate that the understanding of essences that is in the background allows us to state that “Aristotle or the moon is made up of cheese” is essential to Aristotle since that

is a property which would remain invariant through the process of variation in imagination.

Fine is going to distinguish between two approaches to the concept of essence. On the one hand, “an essential property of an object is a constitutive part of the essence of that object if it is not had in virtue of being a consequence of some more basic essential properties of the object; and otherwise, it is a consequential part of the essence” (Fine 1995: 57). So, it is constitutively essential to Aristotle to be human, but it is just consequentially essential to him to be a human, or the moon is made up of cheese.

Contrary to what one should think, Fine claims that we should start from properties that are consequentially essential to somebody or something and then to filter out or deplete this notion to obtain the properties that are constitutively essential.

But how are we to understand the relationship between constitutive and consequentialist essence? One view is that we understand the latter in terms of the former. Roughly, to belong to the consequentialist essence of something is to be a logical consequence of what belongs to the constitutive essence. But another view, to which I am more inclined, is that we understand the former in terms of the latter. One statement of consequentialist essence may be partly grounded in others. The fact that it lies in the nature of a given set to be a set or a set, for example, is partly grounded in the fact that it lies in the nature of the set to be a set. The constitutive claims of essence can then be taken to be those consequentialist statements of essence that are not partly grounded in other such claims. This way of conceiving the distinction enables us to “factor out” the purely essentialist aspect of the concept of essence from the partly explanatory aspect (Fine 2012).

Also, Rosen (2015) proposes something similar when he argues:

p belongs to the constitutive essence of *x* iff *p* to the consequential essence of *x*, and there are no propositions such that *p* belongs to the consequential essence of *x* in virtue of the fact that belongs to the consequential essence (Rosen 2015: 196)

So, Vaidya's method of variation in imagination is problematic because it gives us a consequentialist notion of essence. But what Vaidya means to capture is the constitutive essence. Thus, when we observe through the method of variation in imagination that the property Aristotle is human or the moon is made of cheese remains invariant, we should filter out that claim in order to obtain the property of Aristotle that form parts of his constitutive essence. According to Fine and Rosen, grounding allows us to make the filtering.

However, Eileen S. Nutting, Ben Caplan & Chris Tillman (2017) argue that what they call the "Fine-Rosen proposal" is not an appropriate method, since the result of its application includes in the constitutive essence properties that are not essential, but necessary. Let's consider the reconstruction of the definition of being in the constitutive essence of something proposed Fine that Nutting et al. propose:

The unconstrained Fine–Rosen proposal: For any property *F* and any object *x*, *F* is in *x*'s constitutive essence =df (i) *F* is in *x*'s unconstrained consequential essence, and (ii) it's not the case that there is a property *G* such that the fact that *F* is in *x*'s unconstrained consequential essence is partly grounded in the fact that *G* is in *x*'s unconstrained consequential essence (Nutting et al. 2017: 9).

According to that definition the property "not being Plato" cannot be in Socrates' constitutive essence since (ii) is not fulfilled, for it is the case that there is a property *G*, namely, "Socrates is Socrates" such that the fact that "not being Plato" is in Socrates's unconstrained consequential essence is partly ground in the fact that "Socrates is Socrates" is in Socrates's unconstrained consequential essence.

But this other example is harder to tackle:

(Unconstrained Universal Fact) Socrates's unconstrained consequential essence includes being such that, for any object *x* and any property *F*, either *x* has *F* or it's not the case that *x* has *F* (Nutting et al. 2017: 10)

According to Nutting et al. (2017: 10-11), "unless (Unconstrained Universal Fact) is partly grounded in some fact about Socrate's unconstrained consequential essence, being such that, for any object *x* and any property *F*,

either x has F or it's not the case that x has F will be in Socrates's constitutive essence."

We will argue that (Unconstrained Universal Fact) is just false if the notion of consequence we manage is like the one proposed by Schnieder (2018) a "relevant" one. Schnieder describes the standard notion of logical consequence as the modal notion of consequence: necessarily if the premises in Gamma are true, phi is true: $\Box(\Gamma\alpha\mu\mu\alpha \rightarrow \pi\eta)$. And objects to it that, despite its success, the modal notion of consequence validates principles that are counterintuitive, like, for example, the principle of explosion, and the principle that a logical truth follows from anything. This is precisely the sort of case (Unconstrained Universal Fact) puts forward. So, we explore whether the filtering out of the notion of consequential evidence can be done by substituting relevant consequence for modal consequence.

Consequence is a relation between propositions, so in this case it will be about the relations among propositions concerning essential properties. Schneider (2018) proposes the notion of "web consequence", and he defines it like this:

WC x is a consequence of $y_1, \dots, y_n \leftrightarrow$ under the hypothesis that y_1, \dots, y_n are grounded, one or more of the (thin) grounds of y_1, \dots, y_n (jointly, fully, and thinly) ground x (Schnieder 2018)

According to this definition of consequence, for any object x and any property F , either x has F or it's not the case that x has F wouldn't be in the consequentialist essence of Socrates, for tautologies are not grounded in any other truth (so they are not included in the ground consequence of anything). However, "Aristotle is human or the moon is made of cheese" or "not being Plato" still belong to Aristotle's consequentialist essence under the notion of web consequence.

Thus, Vaidya's method of variation in imagination can be seen as allowing us to identify the constitutive essence in terms of the notion of consequential essence, if the underlying notion of consequence is the notion of web consequence proposed by Schneider, at least it leaves out cases such as the one pointed out by Notting et al.

Starting with the notion of web consequentialist essence, uninteresting cases such as (Unconstrained logical fact) are left out from the start. Then, the filtering goes through the notion of grounding by assessing which properties ground the others and still remain invariant. Aristotle is human or the moon is made of cheese is grounded in the fact that Aristotle is human (a property that remains invariant) for Aristotle. As Hirèche puts it:

Grounding—unlike e.g. mere (strict) implication—is a relation whose relata meet certain *relevance* conditions. More precisely, I will rely on the idea that a full ground should contain *exactly* what is relevant for it to ground what it grounds—i.e. no more and no less than what it needs to do so (Hirèche 2023).

References

- Chalmers, D. (2002) “Does conceivability entail possibility?”, in: T.S. Gendler & J. Hawthorne (Eds.). *Conceivability and possibility*. Oxford: Clarendon Press, 145 – 200.
- Fine, K. (1995) “Senses of Essence”, in: Sinnott-Armstrong, W., Raffman, D. & Asher, N. (Eds.). *Modality, Morality and Belief. Essays in Honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press, 53 – 73.
- Fine, K. (2012) “Guide to ground”, in: Correia, F. & Schnieder, B. (Eds.). *Metaphysical Grounding. Understanding the Structure of Reality*. Cambridge: Cambridge University Press, 77 – 95.
- Hirèche, S. (2023) “Grounding, necessity, and relevance”, *Philosophical Studies*. doi: 10.1007/s11098-023-01968
- Kripke, S. (1971) “Identity and necessity”. In M.K. Munitz (Ed). *Identity and individuation*, New York: NYU Press, 135 – 164.
- Kripke, S. (1980) *Naming and Necessity*. Oxford: Blackwell. .
- Nutting, E.S., Caplan, B. & Tillman, C. (2017) “Constitutive essence and partial grounding”, *Inquiry*, 6 (2): 137 – 161. doi: 10.1080/0020174X.2017.1392895
- Rosen, G. (2015) “Real Definition”, *Analytic Philosophy* 56 (3): 189 – 209. doi: 10.1111/phib.12067
- Schnieder, B. (2018) “On Ground and Consequence”, *Synthese* (198): 1335 – 1363. doi: 10.1007/s11229-018-02012-9
- Vaidya, A. J. (2010) “Understanding and Essence”, *Philosophia* (38): 811 – 833. doi: 10.1007/s11406-010-9243-z

Vaidya, A. J. & Wallner, M. (2018) "The epistemology of modality and the problem of modal epistemic friction", *Synthese* (Suppl 8): 1909 – 1935.

Yablo, S. (1993). "Is conceivability a guide to possibility", *Philosophy and Phenomenological Research*, 53 (1), 1 – 42. doi: 10.1093/acprof:oso/9780199266463.003.0002

Acknowledgement

We would like to thank the Spanish Ministry of Science and Innovation and the Spanish State Research Agency which financially supported this work as part of the research project 'Concepciones deflacionarias en ontología y metaontología' (reference number: PID2020-115482GB-I00). We are also grateful to the Spanish Ministry of Universities, which financially supported this work through the National Program FPU (grant reference: FPU19/00199).

Levels of Abstraction As Family Resemblances From the Classical to the Quantum Mechanical Representation of Reality

Niccolò Covoni (Urbino, Italy)

Silvano Zipoli Caiani (Florence, Italy)

Abstract

This paper explores the challenges in expressing quantum mechanics using natural language and proposes a solution through the application of the Method of Abstraction. Quantum mechanics, deviating from classical physics, presents difficulties in language expression due to phenomena like the double-slit experiment and entanglement. The paper introduces the concept of Levels of Abstraction (LoA) as a framework to analyze information processes. Wittgenstein's ideas on language as a practical tool, shared activity, and the rejection of pictorial representation are invoked. The Method of Abstraction is extended to the quantum domain, resulting in Quantum Levels of Abstraction (QLoA) and Quantum Gradient of Abstraction (QGoA) models. These models provide a structured approach to understanding quantum properties and serve as a bridge between microscopic and macroscopic realms. The revised Method of Abstraction aids in capturing the informational processes in the quantum world, emphasizing the limitations of natural language and the importance of specialized languages, such as mathematics, in comprehending quantum phenomena. The proposed approach aligns with Wittgenstein's paradigm shift in recognizing the role of language rules in understanding different "games" or aspects of reality.

1. Introduction: the two worlds

Quantum mechanics is the field of physics where things happen in a way that is very different to our classical view of the world. Quantum Mechanics is the most accurate description of the world that we have access to. At the same time, however, it is also the most difficult description to understand. This is because the results of quantum mechanics do not seem to fit the categories of our language. Let us elaborate on this slowly.

From the perspective of quantum mechanics, it is possible to notice the limits of the language that we use every day or in the scientific description of the macroscopic world. Some definitions of the natural language are intrinsically different from the ones that we can derive from the theory of quantum mechanics. To give a taste of this difference, we can mention two examples: a famous experiment and a very strange quantum property.

One of the most famous experiments in the history of quantum mechanics is the *double-slit experiment*. In this experiment, we shoot multiple electrons through a double-slit plate in a wall. Without a detector that measures the

electrons passing through the plate, it is impossible to determine where each single electron is passed, and what is only observed is the wave behavior that can be seen in the wall. Instead, with the observation things changed completely: it is possible to detect the slit where each electron is passed, but the interference of the electrons is no more visible. In this example, something very strange for the macroscopic world seems to be happening but it is possible to explain this phenomena with the help of one of the principles of quantum mechanics: the complementary principle proposed by Niels Bohr in 1928.

The principle can be expressed as follows: when dealing with microscopic objects in any experiment, the observer gains insights not into the inherent properties of the objects themselves. Instead, the obtained information pertains to the properties of the objects within a specific context, which includes the use of measuring instruments. Information acquired about the object under particular conditions should be viewed as supplementary to information gathered under alternative conditions. It's crucial to recognize that information obtained in diverse circumstances cannot be merely aggregated or combined to form a unified depiction. Instead, they mirror distinct (complementary) facets of a singular reality, each corresponding to a specific aspect of the object under examination.

This principle seems very different compared to the macroscopic world. However, things are different: the complementary principle applies to every size of matter, while at the macroscopic level it is ignored, in the microscopic world it is not still negligible. Given that, since our language is arranged on a macroscopic scale, it seems unable to account for the parity principle. Differently, the best way to interpret this principle is by using the language of math.

The use of math allows us to understand the situation that is happening during the experiment because it is impossible to measure the two properties when we go under the number at the right of the equation. The world of quantum mechanics becomes very clear if you use mathematics to explain the problems of the lack of direct observation, the use of the Hilbert Space is a great way to understand what seems difficult to catch without the tool of the numbers.

Another typical quantum property that helps to understand the problem of using natural language is entanglement.

Quantum entanglement was first formulated in 1935 by Einstein, Podolsky and Rosen in the famous *EPR paradox*:

"When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, ... By the interaction, the two representatives [the quantum states] have become entangled." (Schrödinger, 1935: 555)

This is a typical quantum property, which has no counterpart in the classical world, but which has several consequences in the theory of the microscopic world; it is impossible to reduce this phenomenon to classical mechanics.

The natural language created by the observation of the macroscopic world is incapable of capturing what this property shows. The only way to talk about it is to use the proper language of physics or, again, the language of mathematics.

If we want to talk about the strange things that happen in quantum mechanics, we need to bridge the specific language of the microscopic world and our natural language as used for the macroscopic world, in doing so the auspicious result is to obtain a correlation between the "two worlds" and a way to reach a better knowledge about quantum phenomena.

The "strange" thing is that at first sight, it seems that the problem lies in the fact that the two systems communicate with each other instantaneously, but this is impossible because it would mean that the communication had traveled faster than light, and according to the principles of special relativity, nothing can break this limit. But this also means that the correlation between the two cannot be explained by the measurement in the classical sense. Entanglement is something deeply different from a classical property. Once again we see the limits of natural language for describing the properties of quantum mechanics.

All the properties of entanglement seem impossible, not only for natural language, but also for a theory of the macroscopic world such as general relativity, which is the best theory of gravity we have. Consider that general

relativity not only improves classical mechanics by extending its physical range (to speeds comparable to those of light and to strong gravitational fields), and not only provides the cosmological model that best describes the evolution of our universe as a whole, but is also the theory whose field equations, once solved, are potentially capable of providing an infinity of cosmological models describing as many physically possible universes.

2. Language as a practical tool

The fundamental inquiry underlying this discourse is the interpretation of the entanglement as a representation of reality. Does it denote something tangible or remain abstract? This age-old question has guided and still guides the debate regarding the ontology underlying quantum mechanics. The cogency of this question grounds in the widely shared assumption according to which every representation that is true corresponds to a matter of fact that makes it true. According to the famous Aristotelian formula, it makes sense to say that a representation is true because it corresponds to facts, while it makes no sense to say that it corresponds to facts because it is true. In this vein, if the quantum entanglement is to be understood as a truthful description of the way the constituents of matter behave, it is assumed that it corresponds to the facts, just as a photograph corresponds to its subject. According to the “pictorial theory of meaning”, a meaningful sentence must share a pictorial form with whatever state of affairs it reports. In this view, the elements of a linguistic representation correspond to elements of the situation they represent, and that the structure of the sentence is shared with that of the situation. However, it is precisely this "pictorial" correspondence that is problematic.

Wittgenstein, in his later work, dismisses his own pictorial representation theory of reality, asserting that the meaning of a proposition lies in usage rather than in the pictorial representation. According to this perspective, the quantum entanglement does not function as a depiction of reality; the crucial aspect is physicists' capacity for calculations, leading to testable predictions. The emphasis is not solely on the measurements, as a positivist might argue, but on the conduct of physicists. The language and mathematics employed serve as tools for regulating and influencing collective human actions to accomplish meaningful work.

Essentially, all the valuable information generated by science exists as forms of the scientific activity, namely, as results of experimentation or calculation. Wittgenstein illustrates this by stating that determining the length of an object involves an activity rather than mere learning of theories and definitions. This perspective implies that understanding quantum physics involves learning how to make scientific activity with it, and vice versa.

Wittgenstein further suggests that mathematics is a shared activity. He poses a hypothetical scenario questioning the belief that “twice two is five,” emphasizing the role of a shared technique that might not be labeled as calculating (RFM I, 168). Accordingly, if we do not perform the correct activity, that is, if we do not use the appropriate set of rules, it is impossible to understand the procedure of mathematical theorems. In this view, mathematics and natural language can be seen as sharing a series of similarities that allow us to consider them as part of the same family resemblance.

As the use of words in language, according to Wittgenstein, also the use of symbols in mathematics is governed by conventions. Following Wittgenstein,, mathematical entities and truths are not discovered, but rather “invented” or created by humans based on conventions that allow shared activities. In other words, mathematical statements and concepts are considered to be human-made activities or agreements rather than reflections of some inherent, objective reality.

Following this suggestion, to inquire about the meaning of quantum entanglement without specifying the corresponding activity – an experiment – is like asking about the sound of a falling tree without a context. Such a question is deemed nonsensical in this philosophical framework.

3. From actions to levels of abstraction

Any language can also be considered as a process of information exchange between people, objects, computers. With this consideration we can take an epistemic structure made in the philosophy of information, which tries to create a way to define all possible processes where there is an exchange of information.

The core definition of this structure is the *Level of Abstraction*.

"A Level of Abstraction, LoA, is a finite but non-empty set of observables. No order is assigned to the observables, which are expected to be the building blocks in a theory characterized by their very definition." (Floridi 2011: 52)

This definition was created by Luciano Floridi and first presented in the *Method of Abstraction* (2004). This structure is capable of analysing any type of information process that can be extracted from a set of observations, from those closest to nature (such as the colour of things) to abstract exchanges of information (prices, analyses of social characteristics).

The main elements of the method are used to define the level at which a system is considered, since each level of abstraction provides a quantified commitment to the type and amount of information that can be extracted from the set under consideration.

An interesting implementation of the method is to analyse the information processes that can be derived from the observables of a given mathematical set, and this can be done with a simple implementation of the elements of the method. Thanks to this we have a method to analyse different sets of observables (like numbers, objects, properties...) and also to make comparisons with them, because we can easily compare the elements of LoA with the well-known rules of set theory, as a result of the definition of LoA as a set.

Another interesting point is that Floridi thinks that his method is not a prerogative of the human species, everyone can use different types of LoA.

"Since they deal with observables, LoAs are not an anthropocentric prerogative but allow a more general (or indeed less biased) approach. We do not have to limit ourselves to human beings or to communities of speakers. Different sorts of empirical or abstract agents, not only human beings but also computers, animals, plants, scientific theories, measurement instruments etc., operate and deal with the world (or, better, with the data they glean from it) at some LoAs. By neatly decoupling LoAs

from the agents that implement or use them, we avoid confusion between CSs, the languages in which they are formulated or embodied, and the agents that use them." (Floridi 2011: 72)

LoAs are not mandatory for each subject, but the possibility of understanding different levels of abstraction is opened up. It is possible to make a comparison with the definition of *family resemblance*:

"I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. – And I shall say: 'games' form a family" (PI 2009: 56).

From this definition each GoA can be seen as a family resemblance and the sum of the LoAs can be seen as a game.

If it is possible to capture the informational processes between mathematical observations, what about the quantum world, which seems to be accessible only through a particular language and a particular mathematics?

The Method of Abstraction can be applied to cases in the world of quantum mechanics, and its explanatory power can be utilized there as well. To achieve this, a promising approach is to link the observables of the method of abstraction with the corresponding concepts in quantum mechanics.

The original version of the Method of Abstraction defines the observable, which can be rearranged using the observable concept in orthodox quantum mechanics. This results in a new observable with more constraints based on physical theory. It is used to define the notion of observable in quantum mechanics as the total energy of a particle with mass m in a real potential field V . To rearrange the Method of Abstraction for quantum mechanics, we must define the observables in that way. So using this definition, various sets of observables can be created according to the rules of quantum mechanics. and these sets can be referred to as observables.

4. The Quantum levels of Abstraction

The *Quantum-Levels of Abstraction* (QLoA) model allows us to map all observables starting from a set of variables and the rules of quantum mechanics. Each QLoA shows us the observables that have the same properties in common. The model maintains the same structure as the original Method of Abstraction, allowing for the creation of multiple QLoAs with different observables. Quantum Levels of Abstraction can be viewed as a sequence of exponentiations of sets. The QLoA that is closer to the observables, less abstract in Floridi's terminology, is the set that contains the observables derived using the self-adjoint operator.

From there, we can take another step and use the notion of *Gradient of Abstraction* (GoA), taken from the original formulation of the Method, to construct similar objects for quantum mechanics.

A GoA is:

"A collection of different LoAs that focus on a particular system or feature forms a gradient of abstraction (GoA)." (Wolf 2012: 24)

The quantum counterpart is a collection of the QLoAs and can be called the *Quantum-Gradient of Abstraction* (QGoA). It can be compared to a quantum state in quantum mechanics. A quantum state is defined as the wave function that encodes all information about a system. The correlation can be identified by observing that each QLoA represents a portion of the information of a specific quantum state, such as position or energy. The set of all QLoAs creates a QGoA that contains all the information of a system. This reformulation gives the QGoA a foundational state compared to the GoA, where this structure is not an axiom of the theory and is intended to aid the analysis of information derived from large collections of LoAs.

By accepting the modification of the Method of Abstraction, it becomes possible to explain the emergence of quantum properties. This is due to the fact that LoAs, which take into account the laws of quantum mechanics, create rules for determining an observable that are not used in a less complex LoA. The latter is used to gain simplicity, such as the LoA that uses the rules of classical physics for macroscopic objects.

The revisitation of the Method of Abstraction can help us understand certain aspects of quantum mechanics that are difficult to express in natural language. It provides a way to bridge the gap between the language we use and the world and language of quantum mechanics. QLoA and QGoA aim to create a bridge between the information processes in the microscopic world and our comprehension.

By reconsidering the concept of linguistic game as a means of analysing the information that can be conveyed, one can comprehend the importance of mathematics as a tool for comprehending phenomena in the microscopic realm, as well as the essential paradigm shift of Bohr's Principle.

Phenomena such as the double slit and entanglement demonstrate the limitations of natural language in capturing precise observations. Therefore, specific tools are necessary to accurately represent reality. The revisited Method of Abstraction enables us to capture both the formal language of mathematics, used to describe quantum phenomena, and the natural language we use for communication, in a common field. This allows for a unified approach, with an high order formal language, to obtaining information from the world, despite the differences between the two theories.

5. Conclusion

Considering these factors, expressing quantum mechanics in natural language may seem impossible. The double-slit experiment and entanglement highlight the difficulty in comprehending the microscopic world without a precise conceptual framework or specific terminology. By incorporating the Method of Abstraction into the quantum realm, we can create a model of the informational processes that serves as a bridge between macroscopic and microscopic situations. This model shows that the limitations of natural language tools derived from the Level of Abstraction that we used and highlights the importance of mathematics and specialised languages in understanding the quantum world. By embracing this paradigm shift, we follow Wittgenstein's ideas of changing the rules of the language in different games and recognising the crucial role of language in comprehending reality.

References

- Floridi, Luciano (2011), *Philosophy of Information*, Oxford: Oxford University Press
- Floridi, Luciano (2004), *Open Problems in the Philosophy of Information, Metaphilosophy*, Oxford: Blackwell Publishing, 554-582
- Schrödinger, Erwin (1935), *Discussion of Probability Relations Between Separated Systems, Proceedings of the Cambridge Philosophical Society* 31, Cambridge, Cambridge University Press: 555-563
- Wittgenstein, Ludwig (1978), *Remarks on Foundations of Mathematics*, G.H. von Wright, R. Rhees and G.E.M. Anscombe (eds.), Oxford: Basil Blackwell
- Wittgenstein, Ludwig (2009), *Philosophical Investigations*, Oxford: Blackwell Publishing
- Wolf, Marty J., Grodzinski, Frances, Miller, Keith W. (2012), *Artificial Agents, Cloud Computing, and Quantum Computing: Applying Floridi's Method of Levels of Abstraction*, In: Demir, H (eds.) *Luciano Floridi's Philosophy of Technology*, Dordrech: Springer, 23-41

A Bipartite Reading of the Private Language Argument of Philosophical Investigations

Haiqiang Dai (Beijing, China)

Abstract

In the “private language” sections (§§243-315) of *Philosophical Investigations*, Ludwig Wittgenstein criticizes the idea of a private language, or a language that can only be understood by its private user and thus is impossible to translate to others. One crucial passage in these sections is §258, which, in a literal reading, shows how a private ostensive definition fails. Commentators disagree upon the features of this passage and hold different views on the private language sections. One group focuses on the memory issue, while the other group emphasizes the reidentification issue. However, I will argue that these two aspects of the private language argument are not in conflict. Both point out that without a public standard of accuracy (a criterion of correctness), private language is impossible. In my memory related interpretation, different interpretations can be reconciled with each other. This essay will proceed in the following way. First, I will explicate and interpret the private language argument. In §258, in order to explain the nonsense of the private ostensive definition, there is a basic line, namely, the lack of a criterion of correctness of memory in the naming ceremony. Then, I will show how this reading could resolve the debate between different interpretations by showing that both the memory scepticism reading and the reidentifications of sensations reading could be interpreted under my reading.

Many commentators regard the §258 of *Philosophical Investigations* as a substantial argument to treat the private language and maintain that this passage is a form of *reductio ad absurdum* in which there is an error in the private ostensive definition such that the possibility of private language is ruled out. As Malcolm says, “The argument that I have been outlining has the form of *reductio ad absurdum*: postulate a 'private' language; then deduce that it is not language”. (Malcolm1954:537) However, this reading diverges into two groups. One group contends that the argument is relevant to the scepticism of memory and argues that the real problem for the private language is the lack of a criterion of correctness. (Canfield2001:379; Hardin 1959:518; Stocker 1966:47) By contrast, the other group believes that the problem is about identifying the sameness of sensation. The failure of the reidentification of sensations at two different times leads to the failure of the private ostensive definition. (Pears 1988:328; Glock 1996:313; Ahmed 2017:50; Lin 2017:273-274)

These readings raise questions that need resolution. First, which interpretation is right with regard to the §258 passage? The criterion related one or the reidentification related one? Textual evidence can support both, it seems. So, is it possible to reconcile them?

In PI §258, Wittgenstein presents a case of a private ostensive definition in which a private diarist attempts to produce a private language through a simple private ostensive definition by concentrating on his sensation and labelling it with a sign. However, this “christening” is not sufficient for the establishment of the private language, as the private diarist believes. Wittgenstein raises a conundrum:

But “I commit it to memory” can only mean: this process brings it about that I remember the connection correctly in the future. But in the present case, I have no criterion of correctness. One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can’t talk about ‘correct’. (PI §258)

The point here seems quite clear: this passage says that the failure of the private ostensive definition is due to the lack of a criterion of correctness for memory. With regard to the interpretation of the content of the criterion of correctness, the text seems to suggest simply the “memory of connection”. However, commentators diverge on this question. It diverges into many types, such as the meaning of the word, the sameness of sensations, and the reidentification of the type of sensation. These interpretations conflict, and they all can find evidence in Wittgenstein’s context. My view is that they are not wrong but that they say the same thing from different perspectives that do not conflict.

All agree: there is no criterion of correctness for the private ostensive definition. Why is this so? To illustrate this problem, let us expand the process of the action of the definition. How many factors are involved in the action? Roughly speaking, there are two-time episodes.

At T_1 , the private diarist has a sensation S_1 and marks a sign S .

At T_2 , the private diarist has a sensation S_2 and also marks the sign S .

For the second mark, he needs three activities. First, he must recall what happens at T_1 , that is, he must recall S_1 and S and the connection between them. Second, he must identify S_1 and S_2 as being the same. Third, he uses the rule established at T_1 to mark S_2 with sign S .

Let us consider the first step, what happens in the recalling of the sensation:

T_1	T_2
S (Sign)	S'
C (Connection)	C'
S_1 (Sensation)	S_1'

Figure 1

As shown in Figure 1, there are three main factors at T_2 :

- (1) S' : the memory of S ;
- (2) C' : the memory of the connection, which is related to the rule;
- (3) S_1' : the memory of S_1 .

For the use of memory, it must be correct. However, in the private case, there is no criterion to confirm the correctness of the memory results. In other words, there is no way to check if one is getting things correct. This is true for the sign, for the connection, and for the sensation.

Memory is unreliable, but this weakness is not the crucial problem. In a significant sense, this feature is that which makes a language sensible. In the *TLP*, Wittgenstein says that the bipolar of right and wrong is fundamental to a sensible language. Rather, the subsequent action with regard to using the memory is what is significant. This relates to checking one's memory. It must be confirmed whether a memory is accurate. How should people check their memories? There are two different situations: for the public case, people could appeal to an outside object that normally would not change; in the private case, there is no such outside object, and the only criterion is a private one. In the latter case, checking privately must rely on another memory. However,

this is problematic. Because memory, in this case, has the intrinsic nature of fallibility, the result of this check is still unreliable. This is because one has merely pushed back or postponed the question of accuracy. In this line, the mortal factor for the private ostensive definition is the problem of infinite regress for the criterion.

This diagnosis has negative effects in many cases:

- (4) The lack of the criterion of the correctness of the memory of S;
- (5) The lack of the criterion of the correctness of the memory of the connection;
- (6) The lack of the criterion of the correctness of the memory of S_1 .

Therefore, in §258, when Wittgenstein says that there is no criterion of correctness, it could mean any of the three things mentioned above. They have the same negative results, though they manifest in different aspects.

Without a criterion of correctness with regard to memory, there is no sample established in the private language case.

In knowing what seeing red is you seem to say to yourself ‘seeing red is this’—you seem to give yourself a sample but you don’t because the usual criteria for the sameness of the sample don’t apply. I can say I call ‘red’ always the same color, or whenever I explain ‘red’ I point to a sample of the same color. (Wittgenstein 1968:236)

The problem of memory has two dimensions: one is that memory lacks the criterion of correctness; the second is that memory cannot be used as a sample. These two dimensions are correlated with each other. For a memory to be used as a sample, there must be criterion for the sameness of the sample. Wittgenstein does not completely deny memory as a mental activity. He simply denies that memory itself can be used as a sample due to the lack of a criterion of correctness. The requirement of a criterion is not an empirical requirement but a semantic requirement. For a proposition to be used, its sense must be determined by a criterion of correctness about the sameness of a sample.

Why can a private diarist not notice that a sample cannot be established in the private language case? The reason is provided by Wittgenstein in the following:

(The temptation to say “I see it like this”, pointing to the same thing for “it” and “this”.) Always get rid of the idea of the private object in this way: assume that it constantly changes, but that you don’t notice the change because your memory constantly deceives you. (PI 2009: 218)

According to this passage, when a person wants to establish a sample of a private object, the real situation could be that the object changes, but memory causes him to believe that there is no such change. This possibility is due to memory’s fallibility and the lack of a criterion of correctness about the sameness of the object. Due to memory’s fallibility, a person may not notice small differences. This causes the person to think that the object is still the same. Without an external (public) standard against which to measure the sameness, there is no opportunity to rectify this incorrect impression.

Many subsequent commentators in the *ad initio* camp argue that the problem is unrelated to memory and insist that the problem is rather about the reidentification of sensation type. In this case, people cannot determine whether S_2 is the same as S_1 . For example, “. . . the argument is that a case can be described in which there would be no distinction between applying a word to a sensation-type correctly and applying it incorrectly.” (Pears 1988:328). While this omits talk of “memory”, it is not incompatible with a reading that emphasizes the role of memory. To see this, it is necessary to examine the reason for the failure of the reidentification in the process of the private ostensive definition.

See the following illustration:

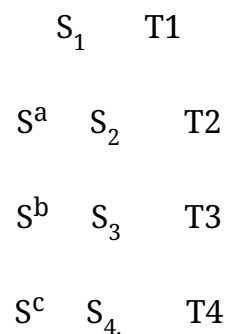


Figure 2

For the whole process, there are two different types of activity. On the one hand, after sensation S_1 occurred at T1, memories of S_1 , S_2 , and S_3 occurred, namely, S^a , S^b , and S^c , at t2, t3, and t4. On the other hand, there are activities of identification, namely, the identification of sameness between (1) S^a and S_2 at T2, (2) S^b and S_3 at T3, and (3) S^c and S_4 at T4. The private diarist thinks that the memories are all correct, so in the first type of activity, he regards S^a , S^b , and S^c as same as S_1 . In the second type of activity, he finds three forms of sameness. With this consideration, he can infer that S_1 , S_2 , S_3 , and S_4 are all the same sensation. However, this final conclusion is not correct because he makes one large mistake. Consider one piece of these series, such as the reidentification between S_1 and S_2 .

For this piece of activity, there are three procedures.

- (A) A memory procedure that establishes the sameness of S_1 and S^a .
- (B) An identification procedure that sets up the sameness of S^a and S_2 .
- (C) An inference procedure that sets up the sameness of S_1 and S_2 .
(Reidentification)

A positive result of (C) depends on the effectiveness of A and B. There is no problem for (B), so the crucial point comes from (A). If the sameness between S_1 and S^a is confirmed, then (C) is valid. However, in (A), the *memory* is fallible, so it needs a criterion to examine the correctness of the sameness. According to §258, there is no such reliable independent criterion available; therefore, (A) cannot be established. This leaves the result that (C) cannot be reached. Similarly, the reidentification of sameness between S_3 and S_4 with S_1 cannot be established.

Therefore, the reason for the failure of the reidentification of the sensation type is that there is no criterion of correctness of the sameness between S_1 and

its *memory* S^a. Along this same line, it is easy to see that reidentifications of aspects of sensation, and of the rule established at the first time, are also correlated with the issue with memory.

As has been shown, commentators disagree about the exact reason for the content of §258. One group focuses on the memory issue, while the other group emphasizes the reidentification issue. However, as I have argued, these two aspects of the private language argument are not in conflict. Both point out that without a public standard of accuracy (a criterion of correctness), private language is impossible. In my memory related interpretation, different interpretations can be reconciled to each other.

References

Ahmed, A. (2017) "Triangulation and the Private Language Argument". In *Belgrade Philosophical Annual* 30, 35-52.

Canfield, J.V. (2001) "Private Language: The Diary Case", *Australasian Journal of Philosophy*, 79, 377.

Glock, H. (1996) *A Wittgenstein Dictionary*, Oxford: Blackwell.

Hardin, C. L. (1959) "Wittgenstein on Private Languages", *The Journal of Philosophy*, Vol. 56, No. 12, 517-528.

Lin, F. Y.(2017) "Wittgenstein's Private Language Investigation", *Philosophical Investigations* 40, 257-281.

Malcolm, N. (1954) "Wittgenstein's Philosophical Investigations", *Philosophical Review* LXII.

Pears D. (1988). *The False Prison*, Vol 2, Oxford: Blackwell.

Stocker, M. A. G. (1966) "Memory and the Private Language Argument", *The Philosophical Quarterly*, Vol. 16, No. 62, 47-53.

Wittgenstein L. (1968) "Notes for Lectures on 'Private Experience' and Sense Data", R. Rhees (ed.), *Philosophical Review* 77, 275–320.

Wittgenstein L. (2009), *Philosophical investigations*. translated by G. E. M Anscombe, P. M. S. Hacker and Joachim Schuller, revised 4th dition by P. M. S. Hacker and Joachim Schuller, Oxford: Blackwell.

Acknowledgement

This work is supported by "the Fundamental Research Funds for the Central Universities" (No. 1233300003).

Ontology Reconstructed: Another Way of Being Neo-Carnapian

Caterina Del Sordo (Donostia-San Sebastian, Spain)

Abstract

To date, neo-Carnapian (meta-)ontological standpoints have focused on the conceptual apparatus of Carnap's *Empiricism, Semantics, and Ontology*, while that of *Der logische Aufbau der Welt* has seldom been explored. This paper addresses this lacuna. In particular, it proposes "ontology reconstructed" (OR) as a neo-Carnapian standpoint based on the antimetaphysical programs of the Vienna Circle's anti-absolutism and of Carnap's early project of rational reconstruction. OR recasts the distinction between internal and external questions of existence and any subsequent deflationism in terms of the distinction between empirical, or relative, questions and metaphysical, or absolute, questions. This paper shows that OR thereby circumvents (1) Quine's objection to the distinction between theoretical-internal and pragmatic-external questions and addresses (2) the neo-Quinean objection of the "duck fallacy", or verbal essentialism. OR is defined via (i) permissivism on existence, (ii) a conception of empirically real entities as entities that satisfy certain structural properties, and (iii) a conception of metaphysically real entities as entities that satisfy no structural properties in principle. OR redesigns the ontological enterprise to identify the properties required by (ii). The paper demonstrates that OR thereby addresses the objections that it (3) changes the subject and (4) engages on uninteresting, or shallow, realism. While addressing (3) and (4), this article outlines a comparison between easy ontology (EO) and OR.

Neo-Carnapian (meta-)ontology occurs in different shapes. Distinctive takes on quantifier variance, fictionalism, the analytic/synthetic distinction, and easy arguments constitute the diversified neo-Carnapian landscape—including Price (2009), Hirsch (2011), Thomasson (2014)—to mention only some. (a) The distinction between internal and external questions of existence and (b) ontological deflationism can be considered hallmarks of neo-Carnapian (meta-)ontology. Moreover, the contemporary neo-Carnapian landscape gathers under the aegis of the (meta-)ontological apparatus outlined in *Empiricism, Semantics, and Ontology* (ESO). Thus, contemporary neo-Carnapians capture only half of the whole Carnapian story. Indeed, as captured in his own words, Carnap recognises that his early arguments on the ontological problems of existence have never been refuted nor critically discussed (see Schilpp 1963: 869).

My first goal in this paper is to identify a neo-Carnapian standpoint, what I call "ontology reconstructed" (OR), that draws upon the (meta-)ontological apparatus of Carnap's early works on rational reconstruction. My second goal is to promote OR by addressing four criticisms of neo-Carnapian

(meta-)ontology, in particular: (1) Quine's objection to the internal/external question distinction, (2) the neo-Quinean objection of the "duck fallacy", or verbal essentialism, (3) the "change of subject" objection and (4) that of uninteresting, or shallow, realism. To achieve these goals, I divide the paper into three main sections. The first outlines the stance of OR on existence, empirical reality, and metaphysical reality by means of Carnap's (1928[1969]) (*Aufbau*) methodological toolkit. The second shows how OR satisfies (a) and (b), thereby addressing (1) and (2), respectively. Finally, the third section addresses (3) and (4) and outlines a comparison between easy ontology (EO) and OR.

1. *Aufbau* on existence, empirical reality, and metaphysical reality

After Quine's (1951[1976]; 1953[1980]) onslaught, there has recently been a Carnapian revival in both the (meta-)ontology and conceptual engineering fields (see, for example, Dutilh Novaes 2020). However, rational reconstruction remains the Cinderella at such a revival.

The reception of Carnap's early works has experienced a sea change in the last 25 years. On the one hand, *Aufbau*'s formal method has been reassessed from a mathematical point of view, in which quasi-analysis appears to be an application of Stone's maxim *avant la lettre* (Mormann 2009; Piazza 1991). Accordingly, rational reconstruction can be seen as a model of conceptual construction (Del Sordo-Mormann 2022: 336-337) or a methodological toolkit. On the other hand, the antimetaphysical spirit of early Neopositivism has been considered an instance of comprehensive anti-absolutism, different from the search for a criterion of cognitive meaningfulness (Parrini 1995[1998]: 18-21). Accordingly, the Vienna Circle's position on antimetaphysics condemned any claim to grasp absolute or transcendental facets of reality as a misunderstanding of the concept of knowledge. It is here, illuminated by the difference between empirical reality and metaphysical reality, that one may rescue Carnap's early (meta-)ontological take.

Metaphysical knowledge is impossible not only because it goes beyond the limit of human reasoning but also because its goals rely on a misunderstanding of the concept of knowledge. [...] Inductive metaphysics, instead, is not logically impossible, to the extent that it leads to testable statements. But, if this is the case, then there is continuity with science,

and the danger simply lies in the fact that one puts forward risky, fanciful, and unfruitful hypotheses. (Feigl 1937: 401, translation and emphasis mine)

One can find the empirical/metaphysical distinction in several Carnapian places, such as *Aufbau* (§§ 170, 175 ff.) and ESO (22). Instead of relying on the distinction between theoretical-internal and pragmatic-external questions, the difference between empirical reality and metaphysical reality depends on the distinction between the absolute and the relative (see also Neuber 2016: 108-110).

In addition to this 'constructional' or 'empirical' problem of reality, the question may arise whether or not we must ascribe 'reality' in a special sense to these empirically real objects. *For this special sense, there are various formulations; most commonly, it is characterized as independence from the cognizing consciousness. Thus, we have to differentiate two different meanings of the word 'reality'.* (*Aufbau*: 283, emphasis mine)

To provide a definition of empirical/metaphysical reality, one appeals to the methodological toolkit of rational reconstruction, particularly the bracketing strategy (*Aufbau*: 101, 274). Rational reconstruction provides a perspicuous representation of a given phenomenon first by bracketing its more abstract and elusive traits and second by topologizing an appropriately chosen set of more concrete and intuitive traits. Mathematical representation employs the bracketing strategy as a forgetful functor of sorts, which momentarily ignores the lattice-theoretical (namely, more abstract) properties of an algebraic structure and takes its order-theoretical (namely, more intuitive) properties as the grounds for reconstruction (Cf. Del Sordo-Mormann 2022: fn. 8; Davey-Priestley 2002: 112). Carnap's early (meta-)ontology employs the same strategy as withholding judgement on reality or unreality and takes existent entities (or even phenomenological essences; see Thomasson 2019: 290) as grounds for reconstruction. Within this context, there is a permissive stance about existence (Cf. *Aufbau*: §1; Schaffer 2009: §2.1):

(i) A candidate entity counts as existent if and only if it fulfils the laws and constraints of a phenomenological domain (be it perceived, experienced, fantasised, imagined, hallucinated, or whatever).

Therefore, if E is the set of existent entities, then E is endowed with an appropriate relation S , giving rise to the structure (E, S) . So, if one takes E as a (meta-)ontological ground and applies Stone's maxim, then the following holds:

(ii) An empirically real entity is an element of the powerset $P(E)$ that satisfies certain structural properties.

From (i), (ii), and the passages quoted above, the following definitions can be agreed upon:

(iii) A metaphysically real entity is not structurally describable in principle.

Evidently, the burden of proof for OR falls on our ability to identify the structural properties required by (ii). The rational reconstruction project has only rough outlines in this respect. However, the contemporary theory of conceptual spaces may provide full-fledged indications. For instance, convex regions of quality dimensions are suitable candidates for structural representations of properties (see Gärdenfors 2014: ch.2). More generally, we let S be a partial order relation on E , giving rise to the structure (E, \leq) . Then, any element of $P(E)$ that is \leq -convex can be considered an empirically real property; otherwise, it is considered unreal. That is, for any element A , $A \in P(E)$ is an empirically real property if and only if for any $x, y \in A$, any element $z \in E$ with $x \leq z \leq y$ belongs to A (other generalisations of the classical Euclidean convexity can be found in Cáceres *et al.* 2005, for instance). *Mutatis mutandis*, the same can be said of *Aufbau's* similarity circles (*Ähnlichkeitskreis*), where similarity is a reflexive, symmetric, and not necessarily transitive relation. Accordingly, any element of $P(E)$, which is a similarity circle, can be considered an empirically real quality; otherwise, it is considered unreal. That is, we let S be a similarity relation on E , giving rise to the structure (E, \sim) . Then, any element $A \in P(E)$ is an empirically real quality if and only if A is a maximal set of similar elements, namely, if $x, y \in A$, then $x \sim y$, and for all $z \in E$, if $z \sim x$ for all $x \in A$, then $z \in A$ (Cf. Mormann 2009: 259).

2. Neo-Carnapian external/internal question distinction and duck fallacy

OR satisfies the hallmarks of neo-Carnapian (meta-)ontological stances. However, unlike ESO, it does not interpret (a) in terms of theoretical and

pragmatic questions but rather in terms of questions about structurally describable entities (see (ii)) and structurally indescribable entities (see (iii)). Similarly, OR interprets (b) by deflating ontology not because its questions have no theoretical point of contention but because their claimed absoluteness violates any form of knowledge (be it pragmatic or theoretical). Therefore, OR may have at least two advantages in addressing (1) and (2).

For (1), the internal/external distinction of OR is unaffected by Quine's (1953[1980]) objection. Indeed, the distinction relies on the separation between structurally describable and indescribable entities, while the objection hinges upon the distinction between theoretical and pragmatic questions. Moreover, OR may appeal to Carnap's early divide between *proper* and *improper* concepts (Mormann 2007: 59-62) to admit the analyticity of some existential statements. That is, the analytic/synthetic distinction of existential statements can be recast by OR in non-rigid terms as practices of justification occurring naturally as different in pure and empirical sciences (Parrini 1995[1998]: ch. 2). Consider, for instance, the different practices used by lattice theory and archaeology to justify existential statements about maximal (or minimal) elements and the ruins of Troy. The statement "for any Boolean lattice, given a proper ideal J , there exists a maximal ideal I such that $J \subseteq I$ " is tested according to the properties of formal abstract, viz. *improper*, relations, while "the ruins of Troy exist" is tested according to the properties of material, concrete, viz. *proper*, ones. In this manner, an analytic/synthetic distinction could have the advantage of not relying upon Carnapian frameworks (see ESO), which would be subject to Van Inwagen's (2020: §4) criticism, or semantic competences, which would be subject to those of Williamson (2007) via Boghossian (1997) (see Thomasson 2014: 235-8; Eklund 2002).

For (2), the neo-Carnapian defence of deflationism is subject to what I generalise as the "duck fallacy", or verbal essentialism, a charge leveled against neo-Carnapians by Van Inwagen (2020: §6). In particular, the duck fallacy arises as soon as one's (meta-)ontological argument cannot be stated without appealing to the old proverb "if it looks like a duck and walks like a duck and quacks like a duck, then it is a duck" whenever seemingly ontologically substantial terms appear (incidentally, this proverb has its own story in Carnap scholarship). Out of metaphor, the problem is one of mistaking already deflated ontological contentions for inflated ones. Due to the

deflationism of OR, one can be more tolerant with philosophical arguments for the existence of entities. As a demonstration, I present two examples of neo-Carnapian deflationism.

Let us consider Thomasson's (2014: §2.5) challenge of across-the-board criteria of existence. According to OR, if an across-the-board criterion concerns empirical entities (see (ii)), then there is no need for philosophers to deflate it. For instance, the Eleatic criterion is structurally characterizable (see Banks 2014: ch.6). Thus, by invoking the internal/external distinction of OR, the Eleatic criterion may count as an internal, deflated, though far-reaching, ontological question. In contrast, versions of the mind-independent criterion may concern metaphysical entities (see (iii)); therefore, they count as external, inflated questions that require philosophers to deflate them. From the point of view of OR, to avoid the duck fallacy in this context, the (meta-)ontological bearing of across-the-board criteria must be evaluated on a case-by-case basis.

Let us now consider Price's (2009) challenge concerning indispensability arguments, or even TYNQUA (following Van Inwagen 2020: §6). Both indispensability arguments and TYNQUA argue for the existence of entities by endowing them with either epistemic relations, such as our best scientific theories for indispensability arguments, or with physico-mathematical relations, such as the ratio of mass to volume in TYNQUA. Therefore, by invoking the internal/external distinction of OR, indispensability arguments and TYNQUA concern internal questions, and hence, there is no need for philosophers to deflate them. I agree with Van Inwagen (2020) that the pivotal point of Price's (2009) challenge is that neither the indispensability argument nor TYNQUA shows the existence of numbers since none of them proceed from ontologically inflated premises. From the point of view of OR, to avoid the duck fallacy, such arguments must be evaluated according to whether they are within or outside the limits of empirical reality (consider (ii)), regardless of how abstract, remote, or even abstruse the claimed existent entities are.

3. Shallow realism and deep realism

Both (3) and (4) raise the charge that neo-Carnapian (meta-)ontology is too deflated. In this respect, the following comparison between EO and OR is not intended to refute OE. More modestly, my aim is to present OR as an additional neo-Carnapian (meta-)ontological standpoint on offer.

According to EO, existential questions are resolvable by simple, or pleonastic, inferences where the premises include undisputed empirical or theoretical claims and the conceptual truth, based on the application conditions of the relevant terms (see Thomasson 2014: ch. 3).

	Inference from undisputed theoretical claims	Inference from undisputed empirical claims
<i>Undisputed claim</i>	4 is even	Mary's healing requires time
<i>Conceptual truth</i>	If P is a property, then $P(x) \leftrightarrow \{x \mid P(x) \text{ is true}\}$ is an abstraction operator.	If P is subject to changes, then P develops
<i>Derivative claim</i>	x is even is a property	Mary's healing is subject to changes
<i>Existential conclusion</i>	A set exists	A process exists

Table 1. Examples of pleonastic inferences.

Objection (3) levels against neo-Carnapian meta-ontology the charge that it initiates a "change of subject" (Cf. Dutilh Novaes 2020: §3.2). Since ontological questions are long-standing, thorny problems, interpreting them in terms of pleonastic inferences is the same as changing the point of contention.

[deflationism] can't give an adequate interpretation of what disputants in serious metaphysics are up to. [...] Neither, it is said, can [one] interpret what serious ontologists are doing in a way that can preserve the idea that there is a real disagreement between those on opposite sides of a metaphysical debate that is non-trivial and worth having. (Thomasson 2017: 1)

Objection (4) levels against neo-Carnapian ontology the charge that it is engaged in only uninteresting, or shallow, realism. Since pleonastic inferences hinge on the way we talk about the world, the reality thereby proven is considered dim, shallow, or uninteresting.

But it is often thought that if we can arrive at ontological conclusions via these trivial inferences, the objects we now say exist can't themselves be very substantial [deep realism]: they must be somehow reduced in ontological standing, mere shadows of language [shallow realism] [...]. (Thomasson 2014: 145)

Propositions and properties are thus 'pleonastic entities,' [...] If, therefore, 'nominalism' entails the non-existence of propositions or properties, nominalism is trivially false. And if 'realism' is by definition true if propositions or properties exist, realism is no more interesting a thesis than is 'If John is married, then John is not a bachelor'. (Van Inwagen 2020: §3)

Ad (3), EO answers that thorny ontological debates can be interpreted as conceptual engineering enterprises or metalinguistic negotiation of the relevant terms. Ad (4), EO answers that its realism is not shallow but simple. Accordingly, the distinction between theoretical and empirical claims can be considered one of degree, depending on the prior theory accepted (see Thomasson 2014: 149). In my view, the answers from EO deserve to be explored in detail. Indeed, although a lengthy discourse on this topic is beyond the scope of this article, EO can explicate the phenomenon of adding new axioms to given frameworks, which Maddy (2013: 69) contends that ESO failed to capture. For example, while constructing an ascending chain, e.g., $x \leq y \leq z \leq \dots$, intuitive and informal conditions arise to apply the term "maximal element". Indeed, the existence of such an element and of a choice function is vigorously debated, making a case for reshaping the contours of mathematics itself (Cf. Davey-Priestley 2002: 52, §10).

Ad (3), OR responds that the objection is misguided. We naturally assume that empirical realism is the OR analogue of simple realism in EO and that OR basically consists of specifying the structural properties required by (ii). Thus, OR is anything but pleonastic, for both empirical science and mathematical

ingenuity enter the field to identify such properties (Cf. Mormann 2009; Gärdenfors 2014). Ad (4), OR responds that the objection is missing the point. Admittedly, one can say that the structural properties required by (ii), such as convex regions or *Aufbau* similarity circles, are ontologically flimsy because they are only formal entities, stripped of meaty, material contents. Nonetheless, OR, via rational reconstruction, is anything but shallow or uninteresting. Indeed, since one can align quasi-analysis with representation theorems (Mormann 2009: 277), OR participates in revealing some "genuine strains of mathematical depth" (Maddy 2013: 81 ff.), to wit: the order-theoretical and topological nature of abstract lattices. Consequently, rational reconstruction can be taken as an instance of Maddy's (2013) normative realism.

4. Conclusion

OR owes its formulation to the renaissance of the neo-Kantian and conventionalist interpretations of Carnap's *Aufbau*, by means of which rational reconstruction has been scientifically reassessed. OR builds on closely related philosophical programs, *i.e.*, the anti-absolutism of the Vienna Circle's antimetaphysics and the *Aufbau* project. Although the project is still at an early stage, OR exhibits philosophical virtues. The external/internal question distinction, when recast in terms of structurally describable/indescribable entities, and (meta-)ontology and deflationism, when recast in terms of the rational reconstruction of empirical reality, enable OR to defeat criticisms such as (1)-(4), which other neo-Carnapian approaches have encountered. Moreover, intriguing questions come knocking at the door of the neo-Carnapian standpoint outlined above: what do the structural properties of empirical reality look like? To what extent does OR contribute to the contemporary debate on quasi-analysis? Finally, what is the relationship between explication and quasi-analysis? This last question is particularly worth debating. To date, the recent discussion of conceptual engineering (see Dutilh Novaes 2020, for instance) has concentrated on the conceptual construction of explication, perhaps overlooking the philosophical potential of quasi-analysis.

References

- Banks, Erik (2014) *The Realistic Empiricism of Mach, James and Russell. Neutral Monism Reconsidered*, Cambridge: Cambridge University Press.
- Boghossian, Paul (1997) "Analyticity", in: Bob Hale and Crispin Wright (eds.) *A Companion to the Philosophy of Language*. Oxford: Blackwell, 331–68.
- Cáceres, José; Márquez, Alberto; Oellermann, Ortrud R.; Puertas, María Luz (2005) "Rebuilding Convex Sets in Graphs", *Discrete Mathematics* 297 (1–3), 26-37.
- Carnap, Rudolf (1928[1969]) *The Logical Construction of the World*, Berkeley and Los Angeles: University of California Press.
- Carnap, Rudolf (1950) "Empiricism, Semantics, and Ontology", *Revue Internationale de Philosophie* 4, 20–40.
- Davey, Brian; Priestley, Hilary (2002) *Introduction to Lattices and Order*, New York: Cambridge University Press.
- Del Sordo, Caterina; Mormann, Thomas (2022) "The Significance of *Quasizerlegung* for Carnap's *Aufbau* and Scientific Philosophy in General", *Philosophical Inquiries* 10(1), 231-253.
- Dutilh Novaes, Catarina (2020) "Carnapian Explication and Ameliorative Analysis: a Systematic Comparison", *Synthese* 197, 1011-1034.
- Eklund, Matti (2002) "Inconsistent Languages", *Philosophy and Phenomenological Research* 64 (2), 251-275.
- Feigl, Herbert (1937) "Moritz Schlick", *Erkenntnis* 7, 393–419.
- Gärdenfors, Peter (2014) *The Geometry of Meaning. Semantics Based on Conceptual Spaces*, Cambridge, MA: The MIT Press.
- Hirsch, Eli (2011) *Quantifier Variance and Realism*, New York: Oxford University Press.
- Maddy, Penelope (2011) *Defending the Axioms. On the Philosophical Foundations of Set Theory*, Oxford: Oxford University Press.
- Mormann, Thomas (2007) "Geometrical Leitmotifs in Carnap's Early Philosophy", in: Michael Friedman and Richard Creath (eds.) *The Cambridge Companion to Carnap*, Cambridge: Cambridge University Press, 43-65.
- Mormann, Thomas (2009) "New Work for Carnap's Quasi-Analysis", *Journal of Philosophical Logic* 38, 249–282.
- Neuber, Matthias (2016) "Carnap's *Aufbau* and the Early Schlick", in: Christian Damböck (ed.) *Influences on the Aufbau*, Vienna Circle Institute Yearbook, 18, 98-113.

Parrini, Paolo (1995[1998]). *Knowledge and Reality*, The Western Ontario Series in Philosophy of Science, 59. Dordrecht: Kluwer Academic Publishers.

Piazza, Mario (1995) "'One Must Always Topologize': Il teorema di Stone, la 'topologia influente' e l'epistemologia matematica", *Rivista di storia della scienza* (ser.II) 4, 1-24.

Price, Huw (2009) "Metaphysics after Carnap: The Ghost Who Walks?", in: David Chalmers, Ryan Wasserman, and David Manley (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press, 320–46.

Quine, Willard Van Orman (1953[1980]) "Two Dogmas of Empiricism", in: Willard Van Orman Quine (ed.) *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 20–46.

Quine, Willard Van Orman (1951[1976]) "On Carnap's Views on Ontology", in: Willard Van Orman Quine (ed.) *The Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press, 203–11.

Schilpp, Arthur (ed.) (1963) *The Philosophy of Rudolf Carnap*, LaSalle: Open Court.

Schaffer, Jonathan (2009) "On What Grounds What", in: David Chalmers, Ryan Wasserman, and David Manley (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press, 347–83.

Thomasson, Amie (2014) *Ontology Made Easy*, New York: Oxford University Press.

Thomasson, Amie (2017) "Metaphysical Disputes and Metalinguistic Negotiation", *Analytic Philosophy* 55, 1-28.

Thomasson, Amie (2019) "What Can Phenomenology Bring to Ontology?", *Res Philosophica* 96(3), 289-306.

Van Inwagen, Peter (2020) "The Neo-Carnapians", *Synthese* 197(1), 7-32.

Williamson, Timothy (2007) *The Philosophy of Philosophy*, Oxford: Blackwell.

The Geach/Black Wittgensteined

Dušan Dožudić (Zagreb, Croatia)

Abstract

Hardly anyone contributed to the growth of interest in Frege's ideas as much as Wittgenstein during some thirty years of his various philosophical engagements. One of the consequences of this impact was the project of translating Frege's philosophical papers, ultimately gathered in the collection *Translations from the Philosophical Writings of Gottlob Frege*, edited by Wittgenstein's students Peter Geach and Max Black in 1952. However, Wittgenstein's influence on the collection editors, Geach and Black, did not stop at his general support for the project or encouragement to study Frege. That he was more directly involved in the project is known for some time now, at least since Geach's 1977 recollection about how Wittgenstein thought of Frege's "Der Gedanke" in 1919 and around 1950 and how that affected his editorial decisions. In this paper, I argue that Wittgenstein had a more profound effect on the matter, which affected the structure and content of the collection more than Geach suggested in 1977. I further argue that this impact on the collection reflects Wittgenstein's philosophical take on Frege's criticism of psychologism (including solipsism and idealism). That, in turn, I take to be a valuable piece of evidence in attempts to understand Wittgenstein's views on the matter that, in one way or another, occupied him at different stages of his philosophical development.

1.

The first book-length English edition of Frege's writings appeared in 1950 (Frege 1950a); the second followed two years later (Geach and Black 1952). In the year between the two editions, Frege's admirer, friend, and critic, Ludwig Wittgenstein, died in Cambridge. One of the last remarks he wrote was about the greatness of Frege's writing style (Wittgenstein 1980: 87), but the remark was merely Wittgenstein's last Frege reference in his four-decades-long philosophical development. There is a history of such references, and they often grew into a promotion of Frege. To start with, Wittgenstein singles Frege out by name, along with Russell, talking in the *Tractatus* preface about "the great works of Frege" to which he owes "in large measure the stimulation of [his] thoughts" (Wittgenstein 1955: 29). The *Tractatus* being a book otherwise lacking references or credits makes that an admirable case of the promotion (given the book's subsequent impact). After the *Tractatus*, Wittgenstein often referred to and reflected on Frege in his manuscripts and lectures, and he promoted Frege in personal communication.

The well-known Wittgenstein recollections of von Wright (1955: 530), Geach (1961: 129-130), and others reveal that Frege was a topic of their conversations with Wittgenstein. The first recalls Wittgenstein "saying that it was Frege's

conceptual realism which made him abandon his earlier idealistic views”; the second one recalls how Wittgenstein described his first encounter with Frege (the famous floor-wiping episode). Malcolm (2001: 70) recalls that during his stay at Ithaca, Wittgenstein sometimes met with him, Max Black, and others to discuss various philosophical issues, including Frege’s “Über Sinn und Bedeutung”. Again, Geach (1991: 14; cf. 1977: viii-ix and 1989: xiii-xiv) recollects: “Shortly before Wittgenstein’s death I often talked to him about Frege; he was pleased at my taking Frege seriously, and gave me much help and advice. [...] The very last time I saw Wittgenstein we were talking about Frege; taking the book in his hands, he said slowly ‘How I envy Frege. I wish I could have written like that.’”

The preserved portion of Frege’s letters to Wittgenstein from 1914–1920 (Frege 2011) and hints about Wittgenstein’s letters from 1913–1920 (Frege 1976: 265-268) testify to the nature of their personal relationship. Being imprisoned as an Austro-Hungarian soldier, Wittgenstein wrote to Engelman from an Italian prison camp asking to send him there a copy of Frege’s *Grundgesetze* (Engelmann 1967: 19). And even after he abruptly terminated his correspondence with Frege in 1920 (cf. Dožudić 2022), he quickly returned to studying his writings: In an October 31, 1920 letter, he asked Engelmann (again) to send him the two volumes of Frege’s *Grundgesetze* (Engelmann 1967: 39). Soon after, he discussed Frege with Ramsey (2012: 143). In 1931, Watson (2012) wrote about Wittgenstein’s advice to dip into Frege’s *Grundgesetze*. Various other sources testify to that relationship, too. Wittgenstein’s manuscripts, scattered remarks, and lecture notes from 1929–1951 reveal much of his later relation to Frege. As Geach pointed out more than once (e.g., 1976; 1977), Wittgenstein often combated with a problem or theme from Frege even when the latter was not explicitly mentioned (cf. Beaney 2017).

All the occupations with Frege manifested during Wittgenstein’s life naturally led his students and friends to take Frege’s ideas more seriously and appreciate how they relate to his ideas. Geach later recalled that he “come to Frege by way of the *Tractatus*”(1991: 16). In her 1959 book on the *Tractatus*, Anscombe explains that “almost all that has been published about [the *Tractatus*] has been wildly irrelevant. If this has had any one cause, that cause has been the neglect of Frege” (1965: 12). It is interesting to note, though, that Wittgenstein’s other student and friend, Norman Malcolm, with whom

Wittgenstein read Frege, managed to write a 14-page encyclopaedia entry on Wittgenstein without even once mentioning Frege (cf. Malcolm 1967).

Despite the early translation of more accessible parts of Frege's *Grundgesetze* in *The Monist* (Frege 1915; 1916; 1917) thanks to Jourdain, but probably also Russell and Wittgenstein (cf. Jourdain 1980: 77; Beaney 2019: 571), the project of translating Frege's writings started in the late 1940s (Frege 1948; 1950a; 1950b; and onwards). By the end of the 1970s, most of Frege's logical and philosophical writings were available in English translation (for the details, see Beaney 2019).

2.

By the 1950s, philosophers close to Wittgenstein, including Geach and Black, came to appreciate Frege and his impact on Wittgenstein, as fittingly reflected in Anscombe (1965). No doubt, much of the appreciation came directly from Wittgenstein. Thus, it is reasonable to assume Wittgenstein affected Geach and Black's decision, if only by paying attention to and praising Frege in their company. Based on Geach and Black's preface to Frege (1952), one might conclude that Wittgenstein's impact on the project was exhausted by such a general reflection and that he contributed nothing in addition significant to the collection. Indeed, after acknowledgements to the journal editors and translators, Geach and Black only add: "Professor Ryle and Lord Russell have been most helpful by lending works of Frege that were otherwise almost unobtainable" (1952: v); not a single word about Wittgenstein. That is puzzling, given Geach's subsequent remarks about the 1952 project and how Wittgenstein actually related to it. In the Preface to Frege's *Logical Investigations*, he (1977: vii) recalls:

[Wittgenstein] took a good deal of interest in the plan Max Black and I had for a little book of Frege translations; and it was through him that I was able to locate some rare works of Frege [...] in the Cambridge University Library.

Later, he affirms and supplements the recollection (1989: xiii-xiv; cf. 1991: 14, 16):

In those last weeks of Wittgenstein's life we saw him several times; [...] He gave me much help and support in the project Max Black and I then had, of translating some works of Frege; he told me where certain articles by Frege were to be found in the Cambridge University Library, and advised me on points of translation. [...] Wittgenstein knew that Russell had preserved in this form [namely, a hard-bound collection of Frege's writings] some articles not easily to be found elsewhere. On Wittgenstein's advice I wrote to Russell, mentioning the source of my information; Russell generously sent the volume round to my house at once.

Wittgenstein, one would think, just because of that, deserved some credit in the collection preface alongside Ryle and Russell. Why he did not get it back then is unclear; Künne (2009: 34) suggests an explanation, but not a particularly appealing one.

Geach further reveals that Wittgenstein did not merely support him and Black in a general and neutral way. From the last quote, that is unclear because advising "on points of translation" sounds like Wittgenstein advised Geach *how* to translate Frege, not *what* to translate. However, Wittgenstein did affect the final selection of Frege's texts more specifically – he advised what to translate – and he did it based on his critical standpoint towards Frege rather than some didactic lines. Geach (1977: vii) reports:

[Wittgenstein] advised me to translate "Die Verneinung", but not "Der Gedanke": that, he considered, was an inferior work – it attacked idealism on its weak side, whereas a worthwhile criticism of idealism would attack it just where it was strongest. Wittgenstein told me he had made this point to Frege in correspondence [...].

Thus, the fact that the first part of the *Logical Investigations* series was left out of the collection is unlikely a coincidence (cf. Floyd 2011: 97; Reck 2002: 27). With the omission, Geach (and Black) either decided to indulge Wittgenstein based solely on his authoritative opinion about the essay or they actually (at the time) agreed with him about the matter. From today's perspective, the call was clearly wrong. Be it as it may, Geach subsequently changed his mind. By 1977, he participated in the translation of Frege's *Investigations* and edited the volume. I take his decision to tell the anecdote with Wittgenstein and "Der

Gedanke” in the Preface as a confession that he (and Black) made a mistake in 1952 under Wittgenstein’s influence. Geach also alluded to the episode in his later recollections (1989; 1991) but never went into such details again.

When the Geach/Black collection was published without “Der Gedanke” and when the essay was independently translated shortly after (Frege 1956), not everybody felt the same about these two events. Dummett wrote two reviews criticising the collection and favouring “Der Gedanke” (1954; 1957), Bennett (1957) a strongly negative one of the essay. Until its English translation appeared, the scholars’ reflections on Frege unaware of the essay already manifested some inadequacies, as cases of Goodstein (1953), Veatch (1954), and Sternfeld (1955) show. The omission of “Der Gedanke” from the easily accessible collection had its price. However, the story of Wittgenstein’s impact on the content of the Geach/Black collection seems not to stop at that: It was more significant than Geach suggested in 1977 and primarily concerned Frege’s distancing from and criticism of psychologism.

3.

Although not evident at first (the way the omission of “Der Gedanke” is), Wittgenstein’s impact on the collection is reflected in Geach and Black’s selection from Frege’s other works. Whether that selection is the result of Wittgenstein’s direct plea for or against a particular text during their conversations about Frege or Geach (and Black) concluded on their own that Wittgenstein’s manifested attitude towards “Der Gedanke” extends beyond it and applies to other texts to be potentially included in the collection is hard to tell. To settle that issue is of lesser importance anyway: On any of the options, it was Wittgenstein who contributed to the outcome. And the fact is that almost all sections in which Frege addresses psychologism, which would be reasonably assumed to be included in the collection, were omitted from it, although the texts of which they are parts often were included. Given the episode with Wittgenstein and “Der Gedanke”, that cannot be a coincidence.

Of the works included in the collection, in “Function and Concept”, “On Concept and Object”, “A Critical Elucidation of some Points in E. Schroeder’s *Algebra der Logik*”, “What is a Function?” and most of the selection from the *Grundgesetze*, Frege does not address psychologism, not even indirectly. As for

the rest of the included works: In *Begriffsschrift* preface, Frege focuses on separating the logical level from the psychological and non-logical, but that consideration is omitted in the collection. In the *Philosophie der Arithmetik*, Husserl provides a psychologistic conception of arithmetic and Frege criticised him primarily on that point. The editors, however, omitted most of that criticism, focusing instead on Frege's additional objections to Husserl. "On Sense and Reference" comes complete, so the reader gets a couple of passages in which Frege separates the objective from the subjective but without directly attacking psychologism. He also briefly touches on the issue of the objectivity of thoughts in "Negation" but assumes the prior discussion in the omitted "Der Gedanke", where psychologism was explicitly criticised.

The case of the *Grundgesetze* foreword selection is puzzling. Compared with "Der Gedanke", its second part significantly resembles the essay's structure and content (cf. Frege 2013: xiv-xxvi; 1977). Nevertheless, Wittgenstein thought the latter to be "an inferior work" that shows Frege did not "grasp deeper grounds of idealism" (Frege 2011: 65). Wittgenstein thought differently about the Foreword, as reflected in McGuinness (1988: 79, 270), who notes Wittgenstein could recite from memory the "glorious" Foreword of his "beloved" *Grundgesetze* (cf. Beaney 2017: 79-80; 2019: 572). What is the deal here? Was Wittgenstein unaware of the similarity? Hardly. Judged by Geach's (1977: vii) recollection, Wittgenstein understood "Der Gedanke" as Frege's *final* attack on "the enemy he had *long* fought", thus, not as something particularly fresh (what, in a way, Frege too acknowledges in a letter (Frege 2011: 47)). Thus, he could not view the essay as an exception compared to Frege's otherwise impeccable opus.

Based on the available documents, I am convinced that what led to Wittgenstein's animosity towards "Der Gedanke" had, in part, nothing to do with its content (Dožudić 2022; similarly, Beaney 2017: 77), and by that time, he already formed the opinion on the *Grundgesetze* foreword). In part, it did, though, and to the extent Wittgenstein disproved Frege's strategy of attacking psychologism, it is also reflected in the selection from the Foreword in the Geach/Black collection: The collection contains only its first part (and the short closure). That part ends precisely when Frege turns to criticism of psychologism.

Geach and Black's decision to omit the second part is unusual. To start with, their collection has the phrase "Philosophical Writings" in the title; yet they omitted the part of the Foreword, which Frege starts by warning readers: "Mathematicians who prefer not to enter into the mazes of philosophy are requested to stop reading the foreword here" (Frege 2013: xiv n. 1). Geach and Black left out this warning without indicating it ever existed. Beaney (2019: 573) suggests an explanation for the omission of the second part: "Although these pages contain some of Frege's most memorable lines (to us now) [...] the focus on Erdmann [whom Frege criticises there] might well have seemed less relevant sixty years later" – at the time the collection was edited. However, I find that explanation problematic.

Frege uses Erdmann as an example to make more general points (Frege 2013: xix), as he often does in his writings. Furthermore, the author's relevance did not govern Geach and Black's other editorial decisions. Frege refers to Thomae in his criticism of formalism as much as to Erdmann in his criticism of psychologism, and Thomae was appealing sixty years later as much as Erdmann, the only difference being that Wittgenstein was fond of Frege's criticism of formalism (Beaney 2019: 574). Indeed, Geach and Black also included texts in which Frege similarly criticised German authors, most of which were also less relevant sixty years later: Kerry in "On Concept and Object", Schröder in "A Critical Elucidation...", Czuber in "What is a Function?", Dedekind, Hankel, Stolz, and Heine (along with Thomae) in the included parts of the *Grundgesetze*. They also included extracts from the Husserl review, the author and (even more so) the book whose relevance for the targeted English readers in the 1950s was dubious. Nevertheless, all the texts ended up in the collection, but the part with the criticism of Erdmann did not. Contrary to Beaney's suggestion, independently of Wittgenstein's interference, it is hard to explain their decision back then as much as it would be hard to explain it today.

If Wittgenstein was equally fond of the whole Foreword, would he advise Geach to omit its part anyway? Would Geach omit it despite Wittgenstein's fondness and put, say, Frege's critical remarks about Schröder instead? I find that unlikely. Given Wittgenstein's interest in the collection and his attitude towards "Der Gedanke", it is more probable that the second part was omitted because of him.

4.

Based on Geach's recollections, one can confidently assume Wittgenstein pleaded for or against Frege's three texts. He was against "Der Gedanke" and favoured "Die Verneinung" and "Über Begriff und Gegenstand" (1977: vii; 1989: xiv). Perhaps he also contrasted some other texts with "Der Gedanke" in the event Geach described or on similar occasions. What readers of the Geach/Black collection may learn about details of Frege's take on psychologism and his refuting strategies is pretty thin. The collection is silent on the matter save several mentioned passages (1952: 59-61, 79, 120-122, 124, 126-127, 146-147), which are indirect attacks on the position at best. It appears Wittgenstein did not think Frege "attacked idealism on its weak side" only in "Der Gedanke"; he must have thought that no text from Frege provides "a worthwhile criticism of idealism".

At this point, the Frege–Wittgenstein relationship with respect to psychologism becomes somewhat complicated. For Frege, various psychologistic, idealistic, and solipsistic conceptions constitute a single untenable subjectivistic class to be treated equally (Frege 2013: xixff.; 1977: 18ff.). Wittgenstein showed antipsychologistic tendencies in the *Tractatus* (1955: 4.1121), but also inclinations towards idealism (Wittgenstein 1979: e.g., 49, 82, 85; 1955: sec. 5.6ff.; Frege 2011: 65, 67). Around that time, he provides Russell with what sounds like a straightforward psychologistic explanation of the Tractarian thoughts (Wittgenstein 2012: 98-99) and evokes a similar strategy in a letter to Frege (2011: 57, 59). In his later writings, he paid some attention to the realism/idealism debate and related matters (e.g., Wittgenstein 1975a: 86; 1975b: 48-49, 178-179). The way and extent to which his take on the matter in the late 1940s and the early 1950s affected Geach and Black's final selection of Frege's writings also contributes to understanding Wittgenstein's thought and its development.

Many influential philosophers promoted Frege's ideas in the first half of the last century before the whole Frege study flourished; the names that first come to mind are Russell and Carnap. It seems, though, that no one contributed to the growth of interest in Frege's ideas as nearly as much as Wittgenstein did during some thirty years of his philosophical engagements. One of the

consequences of these engagements was the project of translating Frege's texts collected in Frege (1952). English translations of Frege's writings have a history, and Beaney (2019: 570) justly observes that "[t]here is a story to tell about each phase in this history". Here, I proposed a chapter on one of the phases and how Wittgenstein figures in it: His influence on the collection editors, Geach and Black, did not stop at the general support for the project or neutral encouragement. Wittgenstein took a more opinionated approach and had a profound effect on the matter that affected the collection's content. Such impact reflects Wittgenstein's take on Frege's criticism of psychologism. That is also a valuable piece of evidence in attempts to figure out his takes on the matters that, in one way or another, occupied him at various stages of his philosophical development.

References

Anscombe, Elisabeth G. M. (1965) *An Introduction to Wittgenstein's Tractatus*, 2nd edition, New York: Harper Torchbooks.

Beaney, Michael (2017) "Wittgenstein and Frege", in: Hans-Johann Glock and John Hyman (eds.) *A Companion to Wittgenstein*, Malden, MA and Oxford: Blackwell Publishing, 74–91.

Beaney, Michael (2019) "A Brief History of English Translations of Frege's Writings", in: Philip A. Ebert and Marcus Rossberg (eds.) *Essays on Frege's Basic Laws of Arithmetic*, Oxford and New York: Oxford University Press, 567–587.

Bennett, Jonathan (1957) "Review of Gottlob Frege: *The Thought: A Logical Inquiry*", *The Journal of Symbolic Logic* 22(4), 392.

Dožudić, Dušan (2022) "Wittgenstein and Frege's 'Der Gedanke': Figuring the Resentment", in: Herbert Hrachovec and Jakub Mácha (eds.) *Platonism: Contributions of the 43rd International Wittgenstein Symposium*, Vol. XXVIII, Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society, 34–36.

Dummett, Michael (1954) "Review of *Translations from the Philosophical Writings of Gottlob Frege*. By Peter Geach and Max Black", *Mind* 63(249), 102–105.

Dummett, Michael (1957) "Frege's 'The Thought'", *Mind* 66(264), 548.

Engelmann, Paul (1967) *Letters from Ludwig Wittgenstein with a Memoir*, Oxford: Basil Blackwell.

Floyd, Juliet (2011b) "The Frege–Wittgenstein Correspondence: Interpretative Themes", in: Enzo De Pellegrin (ed.) *Interactive Wittgenstein: Essays in Memory of Georg Henrik von Wright*, Dordrecht and New York: Springer, 75–107.

- Frege, Gottlob (1948) "Sense and Reference", *The Philosophical Review* 57(3), 209–230.
- Frege, Gottlob (1950a) *The Foundations of Arithmetic*, translated by John L. Austin, Oxford: Basil Blackwell.
- Frege, Gottlob (1950b) "Frege Against the Formalists I–III", *The Philosophical Review* 59(1–3), 79–93, 202 – 220, 332–345.
- Frege, G. (1956) "The Thought: A Logical Inquiry", translated by Anthony M. and Marcelle Quinton, *Mind* LXV(259), 289–311.
- Frege, Gottlob (1976) *Wissenschaftlicher Briefwechsel*, Hamburg: Felix Meiner Verlag.
- Frege, Gottlob (1977) *Logical Investigations*, Oxford: Basil Blackwell
- Frege, Gottlob (2011) "Frege–Wittgenstein Correspondence", in: Enzo De Pellegrin (ed.) *Interactive Wittgenstein: Essays in Memory of Georg Henrik von Wright*, Dordrecht and New York: Springer, 16–67.
- Frege, Gottlob (2013) *Basic Laws of Arithmetic*, Oxford: Oxford University Press.
- Geach, Peter T. (1961) "Frege", in: Gertrude E. M. Anscombe and Peter T. Geach, *Three Philosophers*, Ithaca, NY: Cornell University Press, 126–162.
- Geach, Peter (1976) "Saying and Showing in Frege and Wittgenstein", *Acta Philosophica Fennica* 28(1-3), 54–70.
- Geach, Peter T. (1977) "Preface", in: Gottlob Frege, *Logical Investigations*, Oxford: Basil Blackwell, vii–ix.
- Geach, Peter T. (1989) "Editor's Preface", in: Peter T Geach (ed.) *Wittgenstein's Lectures on Philosophical Psychology 1946–47*, Chicago and London: The University of Chicago Press, xi–xv.
- Geach, Peter T. (1991) "A Philosophical Biography", in: Harry A. Lewis (ed.) *Peter Geach: Philosophical Encounters*, Dordrecht: Springer Science+Business Media, B.V., 1–25.
- Geach, Peter T. and Max Black (eds.) (1952) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Basil Blackwell.
- Goodstein, Reuben L. (1953) "Review of *Translations from the Philosophical Writings of Gottlob Frege*. By Peter Geach and Max Black", *The Mathematical Gazette* 37(320), 141–143.
- Jourdain, Philip E. B. (1980) "Letter to Frege, 15.1.1914", in: Gottfried Gabriel et al. (eds.) *Gottlob Frege: Philosophical and Mathematical Correspondence*, Chicago: The University of Chicago Press) 77–78.

Künne, Wolfgang (2009) "Wittgenstein and Frege's *Logical Investigations*", in: Hans-Johann Glock and John Hyman (eds.) *Wittgenstein and Analytic Philosophy: Essays for P. M. S. Hacker*, Oxford and New York: Oxford University Press, 26–62.

Malcolm, Norman (1967) "Wittgenstein, Ludwig", in: Paul Edwards (ed.) *The Encyclopedia of Philosophy*, Vol. 8, New York: Macmillan Publishing Co., Inc. & The Free Press, 327–340.

Malcolm, Norman (2001) *Ludwig Wittgenstein: A Memoir*, second edition, Oxford: Clarendon Press.

McGuinness, Brian (1988) *Wittgenstein, A Life: Young Ludwig (1889–1921)*, London: Duckworth.

Ramsey, Frank P. (2012) "Letter to Wittgenstein, 11.11.1923", in: Brian McGuinness (ed.) *Wittgenstein in Cambridge: Letters and Document 1911–1951*, Malden, MA and Oxford: Blackwell Publishing, 143–144.

Reck, Erich H. (2002) "Wittgenstein's 'Great Debt' to Frege", in: Erich H. Reck (ed.) *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, Oxford: Oxford University Press, 3–38.

Von Wright, Georg H. (1955) "Ludwig Wittgenstein: A Biographical Sketch", *The Philosophical Review* 64(4), 527–545.

Watson, William H. (2012) "Letter to Wittgenstein, 12.11.1931", in: Brian McGuinness (ed.) *Wittgenstein in Cambridge: Letters and Document 1911–1951*, Malden, MA and Oxford: Blackwell Publishing, 195.

Wittgenstein, Ludwig (1955) *Tractatus Logico-Philosophicus*, translated by C. K. Ogden, London: Routledge & Kegan Paul.

Wittgenstein, Ludwig (1975a) *Philosophical Remarks*, Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1975b) *The Blue and Brown Books: Preliminary Studies for the "Philosophical Investigations"*, Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1979) *Notebooks 1914–1916*, 2nd edition, Oxford: Blackwell.

Wittgenstein, Ludwig (1980) *Culture and Value*, edited by Georg H. von Wright, Chicago: University of Chicago Press.

Wittgenstein, Ludwig (2012) "Letter to Russell, 19.8.1919", in: Brian McGuinness (ed.) *Wittgenstein in Cambridge: Letters and Document 1911–1951*, Malden, MA and Oxford: Blackwell Publishing, 98–99.

Acknowledgement

Parts of this paper were delivered at the "23rd Contemporary Philosophical Issues" conference in Rijeka in 2022 (the 70th anniversary of the Geach/Black collection), where I benefited from the

participants' comments, to whom I am sincerely grateful. The paper was produced within the Zagreb Institute of Philosophy projects *Antipsychologistic Conceptions of Logic and their Reception in Croatian Philosophy* (APsiH, reviewed by the Croatian Ministry of Science and Education and financed through the National Recovery and Resilience Plan by the European Union – NextGenerationEU) and *Intentionality and Modes of Existence* (IP-2022-10-5915, financed by the Croatian Science Foundation).

Luck, Disappearing Agents and Physicalism

Yaren Duvarci (Vienna, Austria)

Abstract

One tendency in the recent free will literature is to argue for a kind of *naturalistic* libertarianism. This is possible by endorsing an event-causal view that takes mental states as either identical with or reducible to brain states and hence positing a libertarian account that is consistent with physicalism. In this paper, I will show two objections directed at event-causal theories. These are luck and disappearing agent objections. I believe that any successful libertarian theory should be able to present a coherent solution to these, since (i) if the outcome of an action is *just* a matter of luck, that action cannot be free and (ii) there is no free will decision without a free agent present. These objections are different, but in my opinion, their solution is the same. I argue that rejecting event-causation and accepting agent-causation solves both problems regarding human action, and agent-causation is needed to have a plausible theory of free acts. Furthermore, I argue that any account of agent-causation is incompatible with physicalism because it denies physical causal closure. I conclude by arguing that physicalism and a strong sense of freedom cannot go together because a physicalist theory cannot reply to these objections.

0. Introduction

One tendency in the recent free will literature is to argue for a kind of *naturalistic* libertarianism. This is possible by endorsing an event-causal view that takes mental states as either identical with or reducible to brain states and hence positing a libertarian account that is consistent with physicalism. In this paper, I will show two objections directed at event-causal theories. These are luck and disappearing agent objections. I believe that any successful libertarian theory should be able to present a coherent solution to these, since (i) if the outcome of an action is *just* a matter of luck, that action cannot be free and (ii) there is no free will decision without a free agent present. These objections are different but in my opinion, their solution is the same. First, I will present the luck objection that is directed against libertarian theories in general. One prominent reply to the luck objection is the reply from control, and I will argue that the kind of control that event-causal theories provide is too weak to defeat the luck objection. Second, I will present the disappearing agent objection and explain why it is a problem for event-causal views. In addition to that, I will argue that what is seen as the rival of event-causation in the libertarianism literature, i.e. agent-causation, can reply to both objections,

and therefore it is the account we should endorse. I will conclude by showing that agent-causation and physicalism are incompatible and no physicalist theory can reply to these objections.

1. Luck

Mele (2006) argues that:

An agent freely A-s at t in a world W requires that it is possible that the agent does not A at t in a world where the past and the laws of nature are the same as W.

The requirement for free will, according to Mele, is that the agent can perform another action than she does in a possible world with the same laws of nature and the same past. But now the problem is since nothing settles whether the agent A-s or not A-s, it is *just* a matter of luck that the agent performs a certain action. Luck objection became a serious issue for the libertarians especially because of what it implies for morality. People think that if we have free will we are morally assessable, but if our actions are just a matter of luck we should not be subject to morality. One prominent reply to luck objection is what is now known as the control principle (or CP). CP states that we are morally responsible for things if they are under our control. If we can show that our actions are under our control, or we have control over them, then we can argue that they are not just a matter of luck and we are responsible for them.

The reason behind this is the intuitive idea if you have control over a situation, that it cannot be *just* a matter of luck. The type of control that we want, however, is another question. Balaguer (2004, 2014), for example, argues that if it is *me* who performs the action, then I have control over the situation. So just because the agent makes the decision, she has control over her actions. The problem here is that, because he embraces the event-causal model, the origin of the action is the psychological events, and not the agent herself since a single event cannot be identified with the agent. This is the reason why the kind of control we are talking about here is limited, that is the kind of control that only can be *passive*. What I mean by passive here is that while the action originates in the agent, the agent is not able to change the course of action

because ultimately what causes the action is the event in her brain that she has no control over in the sense that she cannot change the course of it. Whether the psychological event that causes the action will happen or not is not up to the agent. Moore (2021) gives the example of a leaf drifting down a river to explain passive control. That leaf has passive control over whether it hits the stone due to it being the leaf that hits the stone. However, the leaf itself cannot change the course of action or affect whether it hits the stone or not. So, in virtue of being that very leaf, it has passive control (Moore, 2021, 169). Because the event-causal accounts can give us only passive control, and arguably passive control is not sufficient for us to say that the agent has control over the situation, we say that event-causal accounts cannot provide the kind of strong control we need to defeat luck objection. This is the main reason why the luck objection mainly is a problem for event-causal accounts.

2. Disappearing Agent

Velleman (1992) talks about what he calls the “standard story of human action” in his seminal paper “What Happens When Someone Acts?”. He asks the question that gives the name to his paper, what happens when someone acts, and argues that the standard answer for the question is more or less like this: “His desire for the end, and his belief in the action as a means, justify taking the action, and they jointly cause an intention to take it, which in turn causes the corresponding movements of the agent’s body.” (1992, 461). This is a causal process— the agent’s desires, beliefs, etc. cause the agent to form an intention, and the intention causes the agent’s body to move. This story is mainly credited to Davidson’s *Essays on Actions and Events* (1980) and the causes that make the agent form an intention are seen as psychological and physiological *events*.

Velleman thinks that the standard story is flawed in several ways. He argues that the flaw that concerns him is that the story fails to include an agent, hence the name of the objections on this line “disappearing agent objection”. He argues that:

In this story, reasons cause an intention, and an intention causes bodily movements, but nobody— that is, no person— *does* anything. Psychological and physiological events take place inside a person, but *the*

person serves merely as the arena for those events: he takes no active part.
(1992, 461)

According to the standard story, then, there is a gap between the reasons that are seen as the causes of the action, and the bodily movement itself that is the action. If that is the picture the standard story provides, Velleman concludes, the agent is missing here and the standard story can only capture the deficient instances of agency. He accepts that there are instances of deficient or not full-blooded actions and the standard story can explain those, but if there are full-blooded actions the standard story fails.

I have shown that there is a particular worry that the agent disappears in the standard story of human action. The question, though, is whether this is sufficient to reject the standard event-causal story. Velleman mentions how the defenders of the standard story deem this worry not enough to reject it. So, they think insisting that agents should somehow be in initiation and control of action is a foolish demand and we should not expect the standard story to do that. In Velleman's words, they argue this is like expecting a cake to appear in its own recipe (1992, 462). However, he replies that this is an inadequate response because he thinks the claim that the events recounted in the story don't add up to a person's activity. He says:

Various roles that are actually played by the agent himself in the history of a full-blooded action are not played by anything in the story, or are played by psychological elements whose participation is not equivalent to his.
(ibid.)

He emphasizes that in a full-blooded action, the agents have at least two roles: forming an intention under their reasons and producing behavior that pursues the intention. In the standard story, though, the agent is not doing any of these. He concludes that the agent disappears in the standard story.

3. The Solution

I have mentioned that to provide a sufficient reply to luck objection we need an account that gives the agent a stronger kind of control. This is because for an action to be not just a matter of luck, the agent should be the cause of the

action and be able to change the course of action so that we can say she controls it. Likewise, to provide a sufficient reply to the disappearing objection we need to put the agent back into the causal story somehow. This is possible by either finding an event that can be identified with the agent (which fails for multiple reasons) or positing the agent as a whole as the cause of the action. Structured in this way, both objections boil down to a very similar problem and can be solved by endorsing agent-causation.

The rival of event-causal accounts in causal libertarianist tradition is agent-causation. Agent causal accounts of free will hold that what causes the action are the agents themselves. They typically hold that a form of non-deterministic causation between the intentional/motivational states and the action is not sufficient to provide a strong sense of free will. The reason for that is the lack of agents directly *controlling* the outcome that describes the causal indeterministic scenario. Moreover, there is a genuine causal relation between the agent and the free action (O'Connor, 2011, 4). The distinctive issue here is that agent-causation is understood as fundamentally different from other standard causations, such as event-causation. Event-causal powers are the powers that are disposed to produce effects, whereas agent-causal powers do not produce things, they merely are generic dispositions to cause effects (ibid.) In these accounts, agents have to be taken as enduring and persisting through time; and it is not the mental events that cause the action, but the agents. They are the sole cause and should be present at all times. This gives the agent an *enhanced* control that is needed to defeat the luck objection. Likewise, because the agent causes the action, and not only a single event, the disappearing agent objection is defeated as well. What we need then to have a sufficient reply to both objections is agent-causation. This alone, I believe, provides good reason to endorse agent-causation if we are going to accept some form of causal story regarding human action.

4. Agent-causation and Physicalism

Although agent-causation seemed as obscure and spooky, it gained more popularity in recent years. While Reid (1969) and Chisholm (1976) took human agents as partless, simple substances and argued for a substance dualist ontology, arguably it is not so spooky anymore to argue for higher-level

causation or to argue against reduction. While it is clear a substance dualist conception of agent-causation would not work in the physicalist framework, it is not so clear whether an emergentist agent-causation could do the job considering the growing popularity of non-reductive physicalist theories. In the following, I will explain why agent-causation cannot coexist with physicalism.

Physicalism is the thesis that everything is physical. Seemingly a simple thesis, defining physical has been an incredibly hard job. One can argue for the theory-based conception, object-based conception, or via negativa. One can also be a reductive or non-reductive physicalist, likewise a priori or a posteriori. While there are differences within types of physicalism and how to define it, what physicalists all agree is that the physical realm is causally closed. As Popper says "The physicalist principle of closedness of the physical ... is of decisive importance and I take it as the characteristic principle of physicalism or materialism." (1977, 51). Reductive physicalists accept the thesis by arguing that seemingly non-physical stuff is, in fact, physical and only physical properties enter into causal relations. Likewise, non-reductive physicalists argue that although some properties (like mental properties) cannot be reduced to physical properties, they can only cause things via the underlying physical properties that realize them. For the agent-causationist, though, the situation is different. The agent-causal powers are fundamentally different than the powers of the underlying physical mechanism. Because the agent has novel powers, and because the causal work is done by those powers only the agent has, agent-causation violates physical causal closure while avoiding overdetermination.

There are multiple versions of the closure principle. Kim (2005, 15), for example, formulates it as:

(Closure): If a physical event has a cause at time t , it has a physical cause at t .

Note that this version of the argument does not mention the overdetermination that rules out non-reductive physicalist options for some philosophers. Likewise, it does not overtly argue for the identity of the mental and the physical. Papineau's (2002) argument for causal closure does both:

(P1): If a physical event has a cause at a time t , it has a physical cause at t .

(P2): All mental events have physical effects.

(P3): The physical effects of mental causes are not all causally overdetermined.

(C): Mental events are identical with physical events.

Although Papineau concludes by arguing for reductive physicalism in C, the key part of the causal closure argument is accepting that (i) every physical effect has a physical cause and (ii) there is no overdetermination.

In the previous section, I have mentioned that agent-causation is fundamentally different than other kinds of causation and that it takes agent-causal powers (or agents as a different substance if one wants to go for the substance-dualist route) as fundamentally different than the powers of the underlying mechanism. That is, the causal work in a full-blooded action is done by the agent and the agent is not reducible to the physical thing. It denies physical causal closure openly by arguing that a physical event, like action, has a non-physical cause (since the agent is *over and above* the physical) that cannot be identified with the physical mechanism, and all the causal work is done by the agent. If physicalism is characterized by accepting physical causal closure, and if agent-causation denies this principle; clearly agent-causation and physicalism are inconsistent views.

5. Conclusion

In this paper, I have presented two objections that relate to human action. The first one is directed specifically to libertarian theories. It is the luck objection, and it states that if nothing settles whether the agent A_s or not- A_s , then it is a matter of luck. A reply to the luck objection is saying that if the agent has control over an action, then it cannot be *just* a matter of luck. The problem now is the kind of control that different libertarian theories provide. I have argued that the kind of control event-causal accounts provide is limited and not sufficient to defeat the luck objection. Therefore, what is intended as a problem for all libertarian theories becomes an issue for specifically event-causal accounts. In the next section, I have described the disappearing agent

objection. For a free act, an agent should be present somewhere in the action. Disappearing agent objection emphasizes that while in a full-blooded action, the agent must be somewhere in the story, she is not there because all the work is done by events and the agent merely serves as an arena for these events. Any plausible libertarian account should be able to have a reply to both of these objections. However, this is not possible within the event-causal framework. I argue that if we are going to be libertarians, we should reject event-causation and endorse agent-causation. However, since agent-causation and physicalism are not compatible, a plausible, coherent libertarianism would not be compatible with physicalism as well.

References

- Balaguer, Mark (2004) "A coherent, naturalistic, and plausible formulation of libertarian free will", *Noûs* 36 (3), 379-406.
- Balaguer, Mark (2014) *Free Will*. Cambridge, Massachusetts: MIT Press.
- Chisholm, Roderick M. (1976) *Person and Object: A Metaphysical Study*. Open Court.
- Davidson, D. (1980) *Essays on Actions and Events*. Oxford: Clarendon Press.
- Kim, Jaegwon (2005) *Physicalism, or Something Near Enough*. Princeton University Press.
- Mele, Alfred R. (2006) *Free Will and Luck*. New York, US. Oxford University Press.
- Moore, Dwayne (2021) "Libertarian Free Will and the Physical Indeterminism Luck Objection", *Philosophia* 50 (1), 159-182.
- O'Connor, Timothy (2000) *Persons and Causes: The Metaphysics of Free Will*. New York, US: Oxford University Press USA.
- O'Connor, Timothy (2011) "Agent-Causal Theories of Freedom", in: *The Oxford Handbook of Free Will*, ed. Robert Kane. Oxford: Oxford University Press.
- Papineau, David (2002) *Thinking About Consciousness*. Oxford, GB: Oxford University Press UK.
- Popper and Eccles, Karl (1977) *The Self and its Brain*. New York: Springer.
- Reid, Thomas (1969) *Essays on the Intellectual Powers of Man*. Ed. B. Brody. Cambridge, Mass.: MIT Press. (Originally published in 1785)
- Van Inwagen (2000) "Free Will Remains a Mystery", *Philosophical Perspectives* 14, 1-20.
- Velleman, J. David (1992) "What Happens When Someone Acts?". *Mind* 101 (403), 461-481.

How Neo-Meinongians Take Ontological Questions About Fictional Entities Easy

Manuel Englert (St. Andrews, UK)

Abstract

Amie L. Thomasson developed a deflationist metaontology which encourages us not to worry about ontological questions but to take them easy. Neo-Meinonianism, a competing metaontological alternative, though not often considered from this perspective, can also be described as a straightforward way to settle ontological disputes. I show this using the example of fictional discourse. Thomasson (2015) put forward an argument against a fictionalist treatment of fictional characters. Fictionalists take the ontological claim “There is a fictional character.” not as literally true but only true under pretense or within a game of make-believe. Thomasson pushes against this position by denying that it can be understood what it means to only pretend there is a fictional character, because the pretense cannot be contrasted with a statement where one is really asserting there is a fictional character. The existence of a fictional character trivially follows from an author using a name and pretending to refer to a character. This line of argument, as I show, can be adopted in a Neo-Meinongian framework and likewise used to show that the fictionalists ontological worries are out of place. It will turn out that the Neo-Meinongian approach to ontological question provides us, similar to the deflationist, with easy answers to at least some ontological questions.

Amie L. Thomasson (2015) put forward an argument against a fictionalist treatment of fictional characters. Fictionalists take the ontological claim “There is a fictional character.” not as literally true but only true under pretense or within a game of make-believe. Thomasson pushes against this position by denying that it can be understood what it means to only pretend there is a fictional character, because the pretense cannot be contrasted with a statement where one is really asserting there is a fictional character. In this paper, I am going to show that Thomasson’s argument can be adopted in a Neo-Meinongian framework and likewise used to show that the fictionalists ontological worries are out of place. I proceed in three steps. First, I am going to present three conflicting approaches to fictional entities and their analysis of fictional discourse. Second, I am going to reconstruct Thomasson’s argument against fictionalism, and, lastly, I am going to show how the Neo-Meinongian can adopt this argument and use it against the fictionalist. It will turn out that the Neo-Meinongian approach to ontological question provides us, similar to the deflationist, with easy answers to some ontological questions.

1. Three Approaches to Fictional Entities

There are three approaches to fictional entities relevant in this paper – Artfactualism, Neo-Meinongianism and Fictionalism. (In the context of this paper, when I speak of Neo-Meinongiansism I always mean Modal-Meinongianism as it was developed by Priest (2005) and Berto (2012). One virtue of Modal-Meinongianism is that they account for existence-entailing properties. Unlike nuclear Meinongians, the non-existent object denoted by "Holmes" does not actually have the property of living in Baker Street 221b (Parsons 1980), and unlike for dual copula Meinongians (Zalta 2003) the object does not encode the property, but has the property in the world that realises the Holmes Stories. As a consequence, internal discourse about fictional characters cannot be taken literally but has to be paraphrased.) In a nutshell, they can be characterised by what they claim about the name "Sherlock Holmes". According to the artfactualist, "Holmes" denotes an existent abstract artifact, for the Neo-Meinongian, it denotes a non-existent fictional object and for the fictionalist it denotes nothing at all. These positions also differ in respect to their analysis of internal and external discourse about fiction. As Kit Fine observes, fictional characters come in two ways:

One the one hand, they have certain properties within in the contexts in which they appear, they love and hate, thrive and fail, and live their varied life. On the other hand, they also relate to the real world; they are created by authors, read by readers, and compared, for better or worse, with one and another and with what is real. (Fine 1982: 97)

For the artfactualist, internal statements like "Holmes lived in Baker Street 221b." are true under some kind of pretense, while external statements like "Holmes is a fictional character." can be taken literally. (Thomasson 1998) Neo-Meinongians agree with Realists that internal statements are only true according to a story while external statements can be taken literally. They disagree, however, on the ontological status of the fictional entity. Fictionalists not only hold that internal discourse is only true under some kind of pretense, but also external statements cannot be taken literally. They are like internal discourse true under some kind of pretence or game of make-believe. (Walton 1990, Brock 2002) There are two dimensions to be considered when deciding between these conflicting theories. First, one can evaluate their "ability to give a smoother theory of fictional discourse, and to fit better with our common

sense conception of fictional characters, their creation, existence and identity conditions.” (Thomasson 2015: 2) The analysis of external discourse as true only under pretense or according to a game of make-believe is often seen as heavily revisionary towards fictional discourse, since speakers hold sentences like “Holmes is a fictional character.” typically as straightforwardly true. Fictionalists can justify the revision of fictional discourse by claiming to avoid ontological costs that come along with accepting the existence of fictional characters. This leads to the second dimension of evaluating conflicting theories. One can also consider the plausibility of different ontological commitments of a theory. By accounting for representations without objects of representations, fictionalists seem to reduce ontological commitments, since they are neither committed to the existence of abstract artifacts nor to there being non-existent objects. Should Realists and Neo-Meinongians therefore be worried? Thomasson addresses this worry by combining an artifactualist theory of fictional entities with a deflationary metaontology. She argues not only that there is no problem in claiming that there really are fictional entities, but also that the fictionalist is mistaken that there is anything to worry about. In the next section, we look at an argument that shows why the ontological worry that guides the fictionalists is not justified.

2. Deflationism vs. Fictionalism

Deflationists believe the existence of a fictional character can be derived by an ‘easy inference’ from an uncontroversial claim to an ontological claim. Fictionalists deny that the ontological claim derived through the transformation rules can be understood as a serious claim about the existence of an object. Both, fictionalists and deflationists agree on the truth of the following claim:

- Uncontroversial claim: Jane Austen wrote a novel using the name ‘Emma’ to pretend to refer to and describe a woman.

As we have seen already above, fictionalists and realists both think that statements about fictional characters are true according to a story. When

Austen uses the name 'Emma' she does not refer to a real person, but only pretends to refer to a person. According to Thomasson, the ontological claim can be derived by the following inference:

- **Linking Principle:** If an author writes a story using a name N to pretend to refer to and describe someone, then the author creates a fictional character.
- **Derived claim:** Austen created a fictional character.
- **Ontological claim:** There is a fictional character. (Thomasson 2015: 7)

Given this inference, raising an ontological worry is inappropriate "for the truth of the former guarantees the truth of the latter; nothing more is required." (Thomasson 2015: 7) Yablo (2001, 2005) objects to this kind of inferences. According to him, the linking principle only permits us to conclude that there is a fictional character according to a game of make-believe. When claiming that there is a fictional character, we are only committed to the 'real content' of the uncontroversial claim but not to the one of the ontological claim. Thomasson answers to Yablo's objection by claiming that to make sense of a statement "A pretends that p is the case." one has to be able to explain what it would mean that A commits herself that p really is the case. Sometimes there is an obvious difference between pretending that something is the case and being committed to something being the case. This can be illustrated using an example. If A claims that it is raining today, A is committed to the fact that it is actually raining today. If children play a game within which they agree that it is raining, and they utter the sentence "It is raining today." they are not committed to the fact that it is actually raining, but the sentence is only true within the game. The crux of Thomasson's argument is that this distinction does not make sense in the case of fictional characters. To say one only pretends that there is a fictional character, presupposes an understanding of what it would mean that there really is a fictional character. Since, the claim "There is a fictional character." trivially follows from the uncontroversial claim, there is nothing more that can be said about there really being a fictional character than the uncontroversial claim. (see Thomasson 2015: 15) Consequently, it is not clear about what the fictionalist is actually worrying about when trying to avoid the literal reading of the ontological claim.

3. Neo-Meinongianism vs. Fictionalism

Neo-Meinongians can construct an analagous argument to show why the ontological worry of the fictionalist is mistaken relative to a their metaontological framework. Neo-Meinongians agree with reservations with the uncontroversial claim that Jane Austen wrote a novel using the name 'Emma' to pretend to refer to and describe a woman. As we have seen Emma is a woman only according to a story. According to the Neo-Meinongian, from "Jane Austen represented Emma as a woman." follows that a non-existent object is represented by Jane Austen as a women. What happens when an author creates a fictional character? According to the Neo-Meinongian, Austen imagined Emma, and gave the character she imagined a name. So she was bearing a particular intentional relation to the object, in virtue of which we now imagine Emma. How do we imagine an object? Priest writes:

When Doyle formed the general intention to write about a detective of a certain kind, he had not imagined him; by the time he had written the first story, given him a name, and so on, he had. (Priest 2016: 120)

How exactly an author imagines an object and baptises it, is not an easy matter. Priest (2016) claims this happens through an act of mental pointing. I won't go deeper into that problem, because it concerns the question of the plausibility of Neo-Meinongianism itself. The next step of the inference is the Linking Principle. It states that if an author writes a story using a name N to pretend to refer to and describe someone, then the author creates a fictional character. This may sound problematic for the Neo-Meinongian who holds fictional characters are non-existent objects. And making sense talk of creating fictional characters is, indeed, a contentious issue for Neo-Meinongians. It seems to be a platitude that 'to create' means to bring something into existence. This phrase cannot be taken literally, because fictional characters do not exist. What the Meinongian can say is that Jane Austen was the first to bear that particular intentional relation to Emma. Because she wrote down the story she made it possible that other people can have that intentional relation too (of course they could have had it in principle, this would have required the same imaginative capacities and probably also certain background assumptions they lacked). But this problem aside, the linking principle also holds for Neo-Meinongians. If Austen imagined Emma and writes a story about her, the sentence "Austen created the fictional character Emma." trivially

becomes true at the actual world. From the uncontroversial claim and the derived claim we get: “Austen created a fictional character.” And from that follows the ontological claim: There is a fictional character.

There is a difference in the reading of the ontological claim between the deflationist and the Neo-Meinongian. Deflationists do not distinguish between “there is” and “there exists”. If there is a fictional character, then the fictional character exists. The Neo-Meinongian, however, draws a distinction between the claim that there is a fictional character and that this fictional character exists in the actual world. So, the ontological claim only states that there is an object (which is non-existent) but not that a fictional object actually exists. There is also an interesting commonality between both positions. Both inferences make use of the semantic fact that in order for there to be a fictional character, nothing more – and nothing beyond using the name ‘Emma’ to pretend to and to describe a woman has to be done. “There is a fictional character.” is true when what is stated in the uncontroversial claim is satisfied. Hence, Thomasson’s objection against the fictionalist can be adopted by the Neo-Meinongian. To claim that people are only pretending when uttering that there is a fictional character is besides the point, because that there is a fictional character follows trivially from the uncontroversial claim that an author was bearing a particular intentional relation to an object (i.e. had certain representations) and by naming the object and ascribing it properties created a fictional character. The only way one could meaningfully ask whether there really is a character is to ask whether they exist in our world. If that would be what the fictionalist is asking for, the answer is simply. The object that is denoted by the name ‘Emma’ does not exist.

4. Conclusion and Open Questions

I have shown that a Neo-Meinongian metaontology offers a straightforward response to ontological worries about fictional entities raised by fictionalists. But there remains an open question in the dialectic. Yablo concedes that fictionalism once motivated by avoiding overly ontological commitment has to find a different motivation now:

At one time the rationale for fictionalism was obvious. We had, or thought we had, good philosophical arguments to show that X’s did not exist, or

could not be known about if they did. X's were obnoxious, so we had to find an interpretation of our talk that did not leave us committed to them. That form of argument is dead and gone, it seems to me. It requires very strong premises about the sort of entity that can be known about, or that can plausibly exist; and these premises can always be exposed to ridicule by proposing the numbers themselves as paradigm-case counterexamples. But there is another possible rationale for fictionalism. Just maybe, it gives the most plausible account of the practice. It is not that X's are intolerable, but that when we examine X-language in a calm and unprejudiced way, it turns out to have a whole lot in common with language that is fictional on its face. (Yablo 2001: 87)

To beat fictionalism, Neo-Meinongians not only have to show that they are in a good position to deal with the ontological worry but also that they can give a better account of practice than the fictionalist. This task is to be left open for another occasion. Another open task concerns the generalisation of this account. Deflationism, Neo-Meinongianism and Fictionalism are, mostly, defended across the board. It would be an interesting question how the Neo-Meinongian account fares for example in the realm of mathematical entities compared to fictionalist.

References

- Berto, Francesco (2011) "Modal Meinongianism and Fiction: The Best of Three Worlds", *Philosophical Studies* 152, 3, 313–35.
- Berto, Francesco (2012) *Existence as a Real Property: The Ontology of Meinongianism*, Dordrecht: Synthèse Library, Springer.
- Brock, Stuart (2002) "Fictionalism About Fictional Characters", *Noûs* 36, 1, 1–21.
- Deutsch, Harry (1991) "The Creation Problem", *Topoi* 10, 2, 209–225.
- Fine, Kit. (1982) "The Problem of Non-Existents". *Topoi* 1, numbers 1-2, 97–140.
- Parsons, Terence (1980) *Nonexistent Objects*. New Haven: Yale University Press.
- Priest, Graham (2016) *Towards Non-Being*. Oxford: Oxford University Press.
- Priest, Graham (2005) *Towards Non-Being: The Logic and Metaphysics of Intentionality*. New York: Oxford University Press.
- Thomasson, Amie L. (1998) *Fiction and Metaphysics*. New York: Cambridge University Press.

Thomasson, Amie L. (2015) “Fictional Discourse and Fictionalisms”, in: Stuart Brock and Anthony Everett (ed.) *Fictional Objects*, ed. by Stuart Oxford: Oxford University Press.

Walton, Kendall L. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, Mass.: Harvard University Press.

Yablo, Stephen (2001) “Go Figure: A Path Through Fictionalism”, *Midwest Studies in Philosophy* 25, 1, 72–102.

Yablo, Stephen (2005) “The Myth of the Seven”, in: Mark Eli Kalderon (ed.) *Fictionalism in Metaphysics*, Clarendon Press, 88–115.

Zalta, Edward N. (2003) “Referring to Fictional Characters”, *Dialectica* 57, no. 2.

Acknowledgement

I want to thank Franz Berto for useful discussions and helpful comments.

Imaginary Anthropologies. On Wittgenstein's Last Writings and Epistemic Relativism

Claudio Fabbroni (Berlin, Germany)

Abstract

To Wittgenstein's late thought is often attributed a form of cultural or epistemic relativism, according to which knowledge would always be relative to the criteria of justification valid within a linguistic community. This paper aims to show that this attribution lies largely on a misinterpretation of Wittgenstein's ideas on the relation between language-games and forms of life. In the first section are presented the grounds for some relativist readings of Wittgenstein's thought. In the second section, through the analysis of some passages of the *Tractatus* and *On Certainty*, it is argued that, although Wittgenstein insisted on the "ungroundedness" of our language-games, he did not mean that *any* epistemic attitude, as long as it is culturally endorsed by a community, is as valid as any other one. Rather, it is possible to claim that some games better apply to our world and thus appear as more objective, so that there can be a difference in the validity of world-pictures, contrary to what the epistemic relativist holds. In the third and final section, it is claimed that the different communities that appear in Wittgenstein's examples are not actual or existing alternative possibilities, but rather an imaginary anthropology that Wittgenstein uses to enlighten how *we*, humans, work with *our* language-games. So, it is not possible to attribute him the idea that different games underly different forms of life, as some relativist authors do. It is concluded that Wittgenstein was not likely to be an epistemic relativist.

1. Wittgenstein and relativism

The claim that the "second" Wittgenstein is one of the contemporary fathers of relativism is widely accepted, both by relativist and anti-relativist commentators (e.g. Phillips 1977, Rorty 1979, Arrington, Glock 1996, Baghramian 2004, Boghossian 2006, O'Grady 2015). To him is often attributed a form of *epistemic* or *cultural* relativism, namely the position according to which, since any experienceable aspect of the world is always mediated by enculturation, and thus by the conventions of a given group (Herskovits 1960: 61), there might exist a multiplicity of incompatible and irreconcilable epistemic systems and no single final criterion for adjudicating between them (Baghramian 2004: 67). Hence, knowledge would always be *situated*: it would depend on the culture of reference.

It is not difficult to see why Wittgenstein's late thought could be *prima facie* read accordingly. Indeed, in *On Certainty*, he claims that the foundations of any language-game are certainties, epistemic "hinges" around which the game revolves because if they were removed we would "knock from under [our] feet

the ground” (OC: §492) on which we base *all* our judgments, thus breaking sensible communication (Perissinotto 1991, Moyal-Sharrock 2004, Coliva 2010a, Fabbroni 2023). According to Wittgenstein, some beliefs are “the foundation of operating with thoughts (with language)” (OC: §401), and cannot be themselves proven true nor false, rational nor irrational, justified nor unjustified (OC: §§93–99, 110, 130, 196–206, 222, 307, 499, 559). Certainties are not subject to epistemological investigation, but are “part of our *method* of doubt and enquiry” (OC: §151). That is to say, they provide the grounds on which knowledge can be claimed (OC: §§18, 243). But, indeed, “if the true is what is grounded, then the ground is not *true*, nor yet false” (OC: §205).

Moreover, we do not hold just *some* certainties: language-games are built on a *net* of beliefs, which form “a system, a structure” (OC: §102) that undergoes all our assertions (OC: §162). Indeed, when we learn to play the game “a *totality* of judgments is made plausible to us” (OC: §140); such totality constitutes our world-picture, the scaffolding that “gives our way of looking at things, and our researches, their form” (OC: §211). The world-picture has no verofunctionality as well (OC: §§95, 142, 144): at the bottom of it lies no intellectualization, but “the ungrounded way of acting” (OC: §§105, 110) of the form of life that is playing the game. Now, given that “whether I *know* something depends on whether the evidence backs me up or contradicts me” (OC: §504), and that the criteria of validation for the grounds offered are directly rooted in our world-picture, the fact that a proposition *p* is regarded as true “belongs [...] to the description of the language-game” (OC: §82), because “the *truth* of certain empirical propositions belongs to our frame of reference” (OC: §83).

Wittgenstein’s description of the communal and shared base of the language-games seems very in line with the assumptions of cultural relativism (Glock 1996: 22, 32, 48-50, 110). Underlining the *praxeological, social* character of playing a language-game (Voltolini 1997: 40) supports the idea that truth is relative to some *choice* of criteria of justification. More specifically, it seems to support the claim that the truth of *p* just implies the adherence of *p* to the criteria of truth in a specific system, depending on the world-picture of reference (Marconi 1987: 121-122). Moreover, given that no worldview is in itself justified or unjustified, there could be — and, according to the cultural relativist, there *are* — different epistemic systems, none of which would be intrinsically correct but each of which could certify as justified (and *true*)

different propositions. Indeed, some authors have claimed that Wittgenstein's account might bring to the removal of the possibility of objectivity and knowledge, because "beliefs held within a way of life cannot claim any truth which ought to be accepted by non-participants" (Trigg 1991: 217-218). Similarly, it has been suggested that, given that speaking a language is central to being a form of life (PI §§19, 23, 238, 264), the presence of different games with different underlying world-pictures might imply a difference in *forms of life*. More specifically, "there could be forms of life which use rules of logic and processes of reasoning substantially different from the ones we take for granted" (Baghramian 2004: 76; see also Arrington 1989). This would be reflected in one famous Wittgensteinian example: that of a possible community of people who "sell timber by cubic measure" (RFM: §148). In this case, Wittgenstein wonders,

how could we show them that [...] you don't really buy more wood if you buy a pile covering a bigger area? – I should, for instance, take a pile which was small by their ideas and, by laying logs around, change it into a 'big' one. This might convince them – but perhaps they would say: 'Yes, now it's a lot of wood and costs more' – and that would be the end of the matter. – We should presumably say in this case: they simply do not mean the same by 'a lot of wood' and 'a little wood' as we do; and they have a quite different system of payment from us (RFM: §150).

What he seems to suggest here is that not only a community with such different methods of measurement is imaginable, but also that we might not convince them that the amount of wood has not changed with a change in disposition, because of their very different epistemic methods and viewpoint on reality. Our systems would be simply incompatible and, given the ungroundedness of different language-games, there would be no principled way to say that their beliefs about what is true are incorrect. Indeed, some authors, like Boghossian (2006: 70-72, 108-109), suggest that Wittgenstein would have been sympathetic to the idea that the "cubic-payment community" has a different epistemology or logic from us.

Thus, the epistemic relativist reading of Wittgenstein holds that his late works give rise to an understanding of language-games which implies the fact that different communities, sometimes intended even as different forms of life,

might have incompatible but equally valid ways of intending justifications, truth and even logic. However, the next two sections will argue that this reading lies in a misunderstanding about the ungroundedness of the language-games and the notion of forms of life.

2. Evaluating epistemic systems

Let us return to the claim generally shared by cultural relativists, according to which “where language-games and forms of life as such are concerned no room is left for the notions of truth and falsity” and, therefore, “reason [...] cannot be wrenched apart from [language-games] so as to pass judgment from the standpoint of some contextless and external realm of truth” (Trigg 1991: 215-216). Principally, this argument is not very straightforward: at the very least, it is unclear what a contextless and external realm of truth would be and why would we need it (Vinten 2020: 52-53). Nevertheless, the most interesting point raised by Trigg is the removal of objectivity supposedly made by Wittgenstein that would lead to an infinity of equally but irreconcilable true statements within different games. To give grounds for this claim of *alethic* relativism Trigg quotes a famous passage from the *Investigations*:

‘So you are saying that human agreement decides what is true and what is false?’ — It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life (PI: §241, italics mine).

Nonetheless, on a more careful reading, Wittgenstein states that what is true or false is *not* decided by human agreement. The agreement is in the language-game used, not in the verofunctionality of some statements. Wittgenstein is not claiming that truth is relative to a game or form of life, but he is making a *grammatical* observation of our *use* of the predicate of truth: we assign truth to a proposition when it adheres to the criteria for truth-ascriptions of a certain language-game. However, in a community, something can be deemed true without such a thing making the assertion *ipso facto* true. In fact, the principle that criteria of truth and correctness are only relative to a game and cannot be evaluated from outside of it is quite difficult to defend. Firstly, it leads to counterintuitive conclusions, for why would the statement “Bloodletting will cure your fever” be true for a Greek doctor of the 1st century B.C. and false,

even dangerous, for one of today's EU? But, most importantly, alethic relativism requires ontological relativism, given the need for various, even contrasting and infinite, *true* states of affairs at the same time — a very naïve position never forwarded by Wittgenstein (Vinten 2020: 51-52). Indeed, even if nothing makes a worldview intrinsically better than another, it does not mean that each one has the *same* validity. It is worth noticing that this topic is already introduced in the *Tractatus*, with regards to the example of the description of some irregular black spots on a white surface, which shows how the choice of a set of theoretical principles *is* arbitrary, but *only to a certain extent*. For this purpose, we could cover the surface with a sufficiently fine square mesh, and thus build a “scientific theory” of which true or false propositions (“square *x* is black”, “square *y* is white”) determine a form of world-picture (TLP 6.341). Crucially, the mesh's form

is optional, since I could have achieved the same result by using a net with a triangular or hexagonal mesh. Possibly the use of a triangular mesh would have made the description simpler [...] we could describe the surface more accurately with a coarse triangular mesh than with a fine square mesh (or conversely), and so on. The different nets correspond to different systems for describing the world (TLP 6.341).

So, there is no a priori reason to choose one sort of description over another. In principle, they are all as justified. However, not all possible descriptions are *equal* in terms of accuracy (and simplicity): whether a mesh's point turns out white or black is not an arbitrary circumstance, because it depends on how the spots are distributed on the surface, not on how the mesh is made. Indeed, certain meshes can produce a more precise description of the spots than others. Similarly, although the way of describing the world is arbitrary, the fact that some descriptions are more accurate than others is not, since this depends on *how the world is made* and *not* on how we see it (Frascolla 2006: 128-130). In fact, “we are [...] told something about the world by the fact that it can be described more simply with one system [...] than with another” (TLP 6.342).

Interestingly, Wittgenstein does not seem to have changed opinion in *On Certainty*:

‘But is there then no objective truth? Isn’t it true, or false, that someone has been on the moon?’ If we are thinking within our system, then it is certain that no one has ever been on the moon. Not merely is nothing of the sort ever seriously reported to us by reasonable people, but our whole system of physics forbids us to believe it (OC: §108).

We all believe that it isn’t possible to get to the moon; but there might be people who believe that that is possible and that it sometimes happens. We say: *these people do not know a lot that we know. [...] they are wrong and we know it. If we compare our system of knowledge with theirs then theirs is evidently the poorer one by far* (OC: §286, italics mine).

Given the knowledge and the technology of his time, getting to the Moon was deemed impossible, let alone that someone had already done it. Indeed, if someone believed so, they would just be wrong, they would *know less*. Again, some descriptions, though being *as* arbitrary, apply *better* to the world than others, thus being more objective. In these terms, take the case of the shifts in scientific paradigms as described by Kuhn (1962). At a certain point, it appeared just clear that the Copernican system fitted empirical observations better than the Ptolemaic one, and this led to further discoveries. Even if it was possible, with the addition of quite complicated epicycles, to fit the Ptolemaic one with the observations, the fact that the Copernican system could more simply and effectively explain and predict *does* tell us something about the world: “one can *draw* inferences from a false proposition” (TLP 4.023). In this sense, a community that, for religious or cultural reasons, would still believe in geocentrism, would have beliefs that would not match the actual world and would constitute a poorer description of it than the one that considers modern physics.

Then, world-pictures can change to be more objective, adapting to the world. Moreover, some worldviews adapt better than others, meaning that even if in principle all world-descriptions are justified, not *all* games have the *same* validity. Thus, Wittgenstein just argues about the *ungroundedness* of our language-games, but not their being equally correct in favour of some epistemic relativism.

3. The human form of life

Two questions remain then outstanding: why arguing about this ungroundedness, if not for relativistic reasons? Relatedly, why does Wittgenstein give multiple examples of different communities playing very different games than ours?

About the first question, it is not to forget what is Wittgenstein's main target in *On Certainty*: scepticism; specifically, that kind of sceptical doubt that is meant to be *radical*, questioning *everything* that can be questioned, until the certainty which removes all doubts is reached. Wittgenstein, arguing against this foundational strategy that he tracks back to Descartes (1637/1998) and is exemplified by Moore's (1959) famous "proof of an external world", wants to counter on the one hand the sceptical trope of the infinite regress of the demand for reasons and on the other its obverse, namely that if there is no ultimate, foundational and proven reason, we rely on "assumptions [that are] unguaranteed" (RFM VII: 21). The ungroundedness of our language-games points thus to the fact that there are no ultimate answers: at a certain point, *simply*, when reasons and justifications have been exhausted (OC: §192), "I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do'" (PI: §217). And *this bedrock* constitutes the certainties that govern our game, and which cannot be proven. Therefore, the world-picture is tied to a form of life's way of acting not in a relativistic sense, but in disagreement with the foundational strategy.

Regarding the second question, it is worth noticing that when Wittgenstein, in his last writings, talks about "our" games or worldviews, it is quite unlikely, contrary to what relativists hold, to attribute an empirical sense to this first person plural: "us" is not understood as opposed to any "them" — be it us Westerners, us 20th-century men, etc. Indeed, it is never possible to detect, in the *Investigations* as in *On Certainty*, such a use whereby the plural personal pronoun indicates or refers to a *particular* human group placed in antithesis to another:

Relativism, then, is not really the issue. While the 'we' of Wittgenstein's remarks often looks like the 'we' of our group as contrasted with other human groups, that is basically misleading (Williams 1974: 92).

Rather, when Wittgenstein uses “we” or “our”, he appears to refer to the *human form of life* (Hanfling 2002, Biletzki 2015): there seems to be nothing human that is so different that could *radically* diverge from us. At best it would be a problem of interpretation or translation. Still, there is no human community that is so radically far from us that it could be classified as a *different form of life*. Indeed, the examples of very different communities, like the “cubic-payment” one, seem to have a *self-reflective* aim: “The language-games are rather set up as *objects of comparison* which are meant to throw light on the facts of our language by way not only of similarities, but also of dissimilarities” (PI: §130). In fact, interacting with that community, we may want to revise our translation or understanding of their “a lot/a little of wood” expressions, “because we can’t go against the hinge that one thing — a certain amount of wood — is identical to itself, no matter how we arrange it” (Coliva 2010b: 14). That is, it seems that according to Wittgenstein we should refrain from thinking that they have *radically* different beliefs, because “their knowledge of nature cannot be *fundamentally* different from ours” (RF: 246): the human form of life underlies all possible differences in the games played by different communities, *pace* Baghramian and Boghossian. Coliva (2010b) also underlines this point in analyzing the example of the Azande tribe, which was longly deemed to not perceive logical contradiction (Evans-Pritchard 1937: 23), but *only* due to an error of translation (Bloor 1976/1991: 123-130). Thus, it is precisely because we humans do *not* act in a certain way that Wittgenstein introduces all his different communities. These appear in Wittgenstein’s examples not as actual or existing alternative possibilities, but as an *imaginary anthropology* playing *imaginary games* to cast light on the *actual* games that we play, to bring to light those grammatical structures that define the limits of our world-picture (Andronico 1986: 15).

In conclusion, due to the arguments thus examined, it seems incorrect to attribute epistemic relativism to Wittgenstein: the different communities he portrays are just used to negatively enlighten how *we, humans*, work, and the ungroundedness of our language-games does not mean that any epistemic system is as valid as any other. At best, for this last reason, he can be said to endorse that *very* mild form of relativism that is anti-foundationalism, namely the position to which it is nonsensical to demand ultimate justification criteria for a language-game.

References

- Andronico, Marilena (1986) "Descrivere e immaginare nel secondo Wittgenstein", *Filosofia*, 37(1), 3-44.
- Arrington, Robert L. (1989) *Rationalism, Realism and Relativism*, New York: Cornell University Press.
- Baghramian, Maria (2004) *Relativism*, London: Routledge.
- Biletzki, Anat (2015) "Was Wittgenstein a Cultural Relativist?", in: Danièle Moyal-Sharrock, Volker Munz and Annalisa Coliva (eds.), *Mind, Language and Action. Proceedings of the 36th International Wittgenstein Symposium*, Berlin/Munich/Boston: Walter de Gruyter GmbH, 65-76.
- Bloor, David (1976/1991) *Knowledge and Social Imagery*, London: Routledge.
- Boghossian, Paul (2006) *Fear of Knowledge. Against Relativism and Constructivism*, Oxford: Oxford University Press.
- Coliva, Annalisa (2010a) *Moore and Wittgenstein. Scepticism, Certainty and Common Sense*, Basingstoke: Palgrave Macmillan.
- Coliva, Annalisa (2010b) "Was Wittgenstein an Epistemic Relativist?", *Philosophical Investigations*, 33(1), 1-23.
- Descartes, René (1637/1998) *Discourse on Method and Meditations on First Philosophy*, Indianapolis/Cambridge: Hackett Publishing Company.
- Evans-Pritchard, Edward E. (1937) *Witchcraft, Oracles and Magic Among the Azande*, Oxford: Oxford University Press.
- Fabbroni, Claudio (2023) "Questioning the Body. Certainties between Epistemology and Psychopathologies", in: Ines Skelac and Ante Belić (eds.), *What can be shown cannot be said. Proceedings of the International Ludwig Wittgenstein Symposium, Zagreb, Croatia 2021*, Münster: LIT Verlag, 161-174.
- Frascolla, Pasquale (2007) *Understanding Wittgenstein's Tractatus*, New York: Routledge.
- Glock, Hans-Johann (1996) *Wittgenstein Dictionary*, Oxford: Blackwell.
- Hanfling, Oswald (2002) *Wittgenstein and the Human Form of Life*, London: Routledge.
- Herskovits, Melville Jean (1960) *Man and His Works*, New York: Knopf.
- Kuhn, Thomas S. (1962) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.
- Marconi, Diego (1987) *L'eredità di Wittgenstein*, Roma-Bari: Laterza.

- Moyal-Sharrok, Danièle (2004) *Understanding Wittgenstein's On Certainty*, Basingstoke: Palgrave Macmillan.
- O'Grady, Paul (2015) "Wittgenstein and relativism", in: Maria Baghramian (ed.) *The Many Faces of Relativism*, London: Routledge, 200-222.
- Perissinotto, Luigi (1991) *Logica e immagine del mondo. Studio su Über Gewißheit di L. Wittgenstein*, Milano: Edizioni Angelo Guerini e Associati.
- Phillips, Derek L. (1977) *Wittgenstein and Scientific Knowledge*, London: Palgrave Macmillan.
- Rorty, Richard (1979) *Philosophy and the Mirror of Nature*, Princeton: Princeton University Press.
- Trigg, Roger (1991) "Wittgenstein and social science", *Royal Institute of Philosophy Supplements*, 28, 209-222.
- Vinten, Robert (2020) *Wittgenstein and the Social Sciences: Action, Ideology and Justice*, Cambridge: Anthem Press.
- Voltolini, Alberto (1997) *Guida alla lettura delle Ricerche filosofiche di Wittgenstein*, Roma-Bari: Laterza.
- Williams, Bernard (1974) "Wittgenstein and idealism", in: G.N.A Vesey (ed.), *Understanding Wittgenstein*, London: Palgrave Macmillan, 76-95.
- Wittgenstein, Ludwig (1967) "Remarks on Frazer's *Golden Bough*", Rush Reese (ed.), *Synthese*, 17, 233-253 [RF].
- Wittgenstein, Ludwig (1974) *Tractatus logico-philosophicus*, revised edn., London: Routledge & Kegan Paul [TLP].
- Wittgenstein, Ludwig (1978) *Remarks on the Foundations of Mathematics*, G.H. von Wright, R. Rhees and G.E.M. Anscombe (eds.), revised 3rd edn., Oxford: Blackwell [RFM].
- Wittgenstein, Ludwig (1997) *On Certainty*, G.E.M Anscombe and G.H. von Wright (eds.), amended 1st edn., Oxford: Blackwell [OC].
- Wittgenstein, Ludwig (2009) *Philosophical Investigations*, G.E.M Anscombe, G.H. von Wright and R. Rhees (eds.), revised 4th edn., Oxford: Blackwell [PI].

Acknowledgement

I would like to thank my dear friend Cecilia Brotzu (University of Pavia) for her support and careful feedback on the various versions of this paper.

Alexithymia and Neurodivergence: A Wittgensteinian Deconstruction

Mélissa Fox-Muraton (Clermont-Ferrand, France)

Abstract

This paper examines the pertinence of concept of alexithymia and the assessment tools used to screen for alexithymia (notably the Toronto Alexithymia Scale), especially with regard to the neurodivergent population. Defined as a personality trait involving difficulties identifying and describing feelings and externally orientated thinking, alexithymia is currently thought to affect 55% of Autistics and 41,5% of ADHDers (as opposed to 10% of the general population), and is often referred to as a deficit, impairment or disorder when related to neurodivergent populations. This paper focuses more specifically on autism, and suggests that assessment tools fail to take into account the neurosensory and communication differences in Autistics. It proposes alternative explanations as to why Autistics may score highly on alexithymia evaluations. Drawing on Wittgenstein's reflections from later works on the philosophy of psychology as well as the notion of language-games, it seeks to offer a reframing of the challenges encountered by neurodivergent individuals regarding the communication of feelings and sensations.

While Autism Spectrum Condition has long been associated with difficulties in emotional recognition and reciprocity, new research suggests that these difficulties may in fact represent a specific subset of individuals with co-occurring alexithymia (Kinnaird 2019). More specifically, a recent study suggests 55% of Autistics have some degree of alexithymia (Milosavljevic 2016). The trait is also widely found among other neurodivergent populations, notably 41,5% of ADHDers (Kiraz 2021). Defined in psychiatric literature in the 1970s (Sifneos 1973), alexithymia is technically not a diagnosable condition, rather a personality trait concerning 10% of the general population. However, alexithymia is often referred to as a 'disorder' presenting pathological 'symptoms' or 'impairment,' and more popularly as 'emotion-blindness' or a 'world without emotions,' especially in contexts where this trait is associated with persons with disabilities. The consensus in research today defines alexithymia as a multidimensional personality trait involving difficulties identifying one's feelings, difficulties describing feeling, and externally orientated thinking (Preece et al 2020). While some research does show differences in brain activity, alexithymia is most often diagnosed using self-reports through verbal assessments, where individuals are asked to respond to a series of questions about their ability to understand and express their emotional states. The most-used tool for this is the Toronto Alexithymia Scale, a twenty-question Likert-scale test providing possible response choices ranging

from strongly disagree to strongly agree. This tool includes statements such as: “It is difficult for me to find the right words for my feelings”; “I have physical sensations that even doctors don’t understand”; “I prefer to analyse problems rather than just describe them”; “I have feelings that I can’t quite identify”; “Being in touch with emotions is essential.”

In this paper, I aim to offer a Wittgensteinian exploration of the problems related to the use of such tools and concepts regarding the neurodivergent, and more specifically the Autistic, population (although many of the arguments may also apply to other forms of neurodivergence). This is particularly important insofar as alexithymia assessments are becoming more widespread in autism evaluations, and impacting the understanding of profiles and support needs (Ferguson 2023). While both in research and within neurodivergent communities it has become common to speak of alexithymia, the notion of alexithymia may however not do justice to the particular neurosensory (Daniel 2023) and communication differences of neurodivergent individuals. Notably, the concept of alexithymia fails to take into account the fact that emotions are socially constructed concepts (Feldman Barrett 2017) that rely heavily on the language-games through which we learn their how to use expressions of sensation or feeling. Neurodivergent individuals, however, may have both different sensory experiences of the world, and different ways of processing these experiences, leading to differences in the ways in which they play language-games. As Wittgenstein points out, we are all *taught* how to use expressions of feeling in meaningful ways within the context of shared language games—that is, we are taught what constitute understandable expressions and situations in which these should arise (Wittgenstein 2001, §244). There are ‘appropriate’ contexts in which it makes sense in our language-games to use expressions such as ‘I am in pain’ or ‘I am sad.’ The death of a loved one would constitute a generally recognized acceptable circumstance for the latter, where both the linguistic expression, and other non-verbalised expressions of grief, such as crying, would be perceived as comprehensible and expected.

Problems arise, however, when our language-games do not match up. What happens if a person says ‘I am sad’ or begins to cry after losing a pencil? Or if a person fails to cry at the funeral of a loved one, and *claims* not to *feel* sadness? In these contexts, one might assume that the person does not understand the

concept of sadness and the way in which they are supposed to be playing the language-game. Alternatively, the person may be seen as demonstrating a pathological reaction to the situation, an expression of inappropriate emotions.

Both the pathologizing view and the lack of understanding view fail to take into account, however, that there may be substantial differences in the ways in which individuals experience the world, process information, and communicate verbally and non-verbally. They presume either a lack of competence or a lack of emotional intelligence/regulation on the part of the individual, rather than taking into account the specificities of the ways in which neurosensory differences impact both experience and expression for neurodivergent individuals. These models deny neurodivergent experiences and communication, rather than exploring the possibility that neurodivergent individuals may be playing very different language-games than those used in neuronormative settings. They presume that there is a single, ‘right’ way to relate to and express events and emotional states. This can be particularly detrimental to neurodivergent individuals, especially regarding suggestions to tailor ‘therapeutic’ approaches to neurodivergents based on recognizing and mimicking ‘typical’ emotional expression, as this could further contribute to alienating individuals from the reality of their lived experience, forcing them to ‘mask’ rather than to understand their own experiences and advocate for their needs.

While the concept of alexithymia was initially described in relation to patients with diagnosed mental health conditions in whom observable, measurable physiological responses were attested, while patients reported not feeling anything, it is now used regarding a broad range of very different contexts. These include: 1) absence of reported or observable emotive responses; 2) individuals who dislike *speaking* about emotions or have difficulty doing so using verbal language; 3) individuals who tend to describe their experiences in factual terms rather than use emotion-based vocabulary; 4) people who report difficulty clearly identifying their emotional states (such as determining, when one is upset, whether one is sad, angry, scared); 5) people who have difficulty identifying the *cause* of their emotional states (i.e.: determining why one is angry or sad). Is alexithymia then a difficulty in putting words onto something ‘inner’ that is really there, and if so is this the result of an impairment or difference within the individual (such as difficulties with interoception or

metacognition)? Alternatively, does alexithymia result from real differences in the ways in which emotion or sensation are experienced and expressed, leading to a difficulty playing specific expected language-games? The concept of alexithymia, as it is currently used in psychology, seems to presuppose that there are ‘inner’ states that *should* be *universally* experienced, and imposes a deficit-based model of understanding on individuals who do not demonstrate or relate to expected expressions, behaviours and contexts of feeling. It assumes that feelings and sensations are all experienced and expressed in a universal way, without taking into account individual variation.

Wittgenstein’s late writings on psychology evoke issues of how we should interpret differences in perception and communication. He asks what notions such as blindness and seeing can mean to different individuals: “I can teach human beings the meaning of the words ‘blind’ and ‘sighted’, and indeed the sighted learn them, just as the blind do. Then do the blind know what it is like to see? But do the sighted know?” (Wittgenstein 1992, 75). Wittgenstein suggests that it is not an issue of learning concepts, but rather one of life: “the lives of the blind are different from those of the sighted” (1992, 75). One may very well be able to understand concept of ‘sight,’ but this does not entail that these will be used in the same ways, or even in the same language-games, for individuals navigating the world through different experiences.

The challenge that this poses, as with the case of alexithymia, suggests that it is not so much a problem of a person’s inner experiences or mental states, but rather one of communication. Wittgenstein notes that regarding expression of feeling, the difficulty arises when expressions do not concord with expectations: “Someone smiles and his further reactions fit neither a genuine nor a simulated joy. We might say “I don’t know my way around with him. It is neither the picture (pattern) of genuine nor of pretended joy.” Mightn’t his relation to a person with normal feelings be like that of a colour-blind person to the normal-sighted?” (1992, 61). Expressing feeling in such a way that the other does not know how to interpret the expression (in this case, through a smile) leads to *communicative* difficulties—we do not know how to react, how to interpret this particular behaviour. We might assume indeed that it is a type of emotion-blindness, a different or impaired way of experiencing and interpreting the world. However, Wittgenstein is cautious about this type of interpretation. As he notes, when we “can’t figure...people out,” we should

perhaps rather question our own belief that we can and should always be able to determine someone's mental state, and instead understand that what is posing a problem for us is our own set of beliefs: "Isn't it their reactions that I can't figure out? That for instance I cannot foresee; that keep on surprising me?" (Wittgenstein 1992, 69).

I would suggest that there are a number of reasons for which neurodivergent and especially Autistic individuals (like myself) may have high scores on alexithymia tests, which do not reflect the individual's actual ability to accurately identify their own feelings and sensations. Importantly, Autistic people do have different neurosensory profiles, which means that we experience the world differently to non-autistics: the touch of a certain fabric, or a particular sound or colour might be experienced as intensely painful, or to the contrary as a source of intense pleasure, where to non-autistic individuals these are just more or less neutral aspects of the environment. From birth, the Autistic person's reactions (cries of pain or expressions of joy) are however interpreted through the lens of non-autistic experience—a particular expression will be greeted by those around them with the normalizing aim of instructing the infant or child in how to (re)act or interpret certain situations, very often through negative and invalidating feedback—disregarding and shaming autistic reactions as over-reactions, being dramatic or oversensitive, acting out, making things up for attention, etc. Autistic individuals may be repeatedly informed that our crying or laughter is inappropriate, that our expressions of pain are unfounded, that our non-verbal expressions of emotion (facial expressions, posture, etc.) are inappropriate for the situation. As Wittgenstein notes, the "belief that our concepts are the only reasonable ones consist[s] in [the fact t]hat it doesn't occur to us that others are *concerned* with completely different things, and that our concepts are connected with what interests us, with what matters to us. But in addition, our interest is connected with particular facts in the outer world" (1992, 46).

Every Autistic person has their own unique sensory profile, yet neurosensory differences are an integral part of autistic experience. The difficulties we encounter with communication, especially around feeling and sensation, may to a great extent be linked to the lack of understanding on the part of those around us, and of specific language in which to express our experiences. This is not to claim that our experiences are private sensations that would only be

expressible through a private language; rather, that the shared public language of non-autistic individuals (neuronormative language), relating to the ways in which those individuals experience the world and the objects in it that they are concerned with, may not provide us with the concepts and means of expression to adequately convey our experiences. Moreover, the spontaneous reactions that we may have to certain situations (e.g. being in a noisy environment, being forced to sit still, etc.) may lead to expressions which are interpreted by others as irritation, anger, disinterest, boredom, apathy, etc., when in fact they are merely a response to sensory stimuli. Without the knowledge that our experiences are different to those of others, or other means of interpreting the situation, we will likely take on the concepts that others project upon us and our outward behaviours. Unlike other types of perceptual differences, such as colour-blindness, which can be tested for to ascertain whether disagreement about colour is the result of such a difference (Wittgenstein 1992, 24), there are no specific tests to establish sensory processing differences, and communication is particularly difficult for non-speaking or pre-speaking individuals. If my lack of eye-contact is described by others as disinterest or boredom, or my sensory discomfort interpreted as irritation or anger, I may very well end up with a very different contextual understanding of the use that these terms play in language-games. If I am instructed to smile and say I am happy despite being in pain and distress, because it is socially acceptable or expected, how am I supposed to learn to use expressions in ways that correlate to my experience? Even if I do recognize the dissonance, I may have no specific language in which to communicate my experiences.

Both neurosensory differences and the process through which we learn feeling and sensation concepts can provide important clues as to why Autistic individuals might struggle with expression of feeling, and especially why we might respond in particular ways to the types of questions asked on alexithymia questionnaires. Hyper- or hyposensitivity can certainly lead one to have multiple experiences of physical sensations that doctors do not understand, especially if one does not have the knowledge or ability to communicate that these are common autistic experiences. Difficulties finding the right words to express feelings can arise from years or decades of being told that one's spontaneous expressions of pain or joy are inappropriate for a

given situation. The constant invalidation one might encounter, as well as an obligation to mask one's spontaneous reactions, could also easily lead a person to suppress feelings, or to consider them inessential.

It is not merely neurosensory differences that need to be taken into account, however. Autistic individuals also communicate differently, generally with a tendency to favour more direct, information-based communication. Wittgenstein reminds us that “we not only fail to understand someone else when he hides his feelings, but frequently also when he does not hide them, indeed when he does his utmost to make himself understood” (1992, 28). And here again, it is a divergence in language-games that may provide an explanation. Although neuronormative language-games are based on following socially acceptable scripts, such as demonstrating and expressing sadness at a funeral, an Autistic person might not follow the social script, and might give an honest response. When asked if they ‘feel sad,’ they may very well say no, because that would not be an adequate description of their state at that time: perhaps what they are feeling at that precise moment is numb, or heavy, or lethargic. Perhaps they are experiencing sensory overwhelm due to the lighting and echoes in the church, and *feel* itchy and restless. While this type of response challenges the neuronormative language-game, it should not be necessarily interpreted as indicative of an inability to identify emotion and even less a lack of empathy. Rather than dismiss such responses as a lack of emotional awareness, should we not rather assume that perchance “exactly the same thing is going on within him as within me, only it is expressed differently?” (Wittgenstein 1992, 28).

Communication differences can also explain why Autistic individuals may have difficulty responding to direct questions about feeling and emotion, as these questions are often framed in broad, general terms: i.e., ‘how are you doing/feeling?’ While these types of open questions appeal to non-autistics, Autistic individuals tend to struggle with these questions, and a typical (although not always articulated) answer can be ‘I don’t know,’ or ‘What aspect of my life are you enquiring about?’ Theories about autism developed by Autistic researchers offer convincing arguments as to why this may be the case. “Monotropism” (Murray et al. 2005), the first theory of autism developed by Autistic researchers, convincingly suggests that the distinguishing feature of (at least some) autistic profiles is the way in which attention is focused on a

single task or topic at a given time, as opposed to the more frequent split attention among the general population. While a monotropic mind is able to focus more intensely, this also entails difficulties with transitions, or changing contextual situations. When asked to respond to questions about how one feels, a monotropic thinker might have difficulty getting out of their immediate point of focus, leading to seemingly incongruous remarks from the perspective of the person asking the question, such as a monologue on the person's current topic of interest, or a remark on the sensory experience they are currently focused on (e.g.: 'it's too hot in this room'). While from the perspective of a non-autistic person these might appear to be evasive responses or to convey an inability to get in touch with emotions, from the perspective of the Autistic person they may be very accurate depictions of their current state, and be communicating the information they wish to communicate. As Damian Milton has argued, conceptions of Autistic individuals as incapable of empathy or unable to understand emotions relied on judging Autistic communication with neurotypicals; Autistics are however no worse at empathizing and communicating with other Autistics than neurotypicals are with other neurotypicals. Problems occur in cross-neurotype communication; this is now commonly referred to as the "double-empathy problem" (Milton 2012).

I am of course not claiming that alexithymia cannot be a helpful term for certain individuals in clarifying their challenges and finding solutions. My contention is, however, that evaluating neurodivergent individuals' understanding of and expression of our emotional states and feelings based on neuronormative patterns can undermine the reality of neurodivergent experience. Rather than require that neurodivergent individuals conform to neuronormative expectations, we should learn to expand our language-games, and recognize with Wittgenstein that indetermination and variability are the foundations of both human life and language (1980, §626-627). As Wittgenstein notes: "the main difficulty arises from our imagining the experience (the pain, for instance) as a thing, for which of course we have a name and whose concept is therefore quite easy to grasp...What we don't see is that the *concept* pain is only beginning to be investigated" (1992, 43).

References

Daniel, Helen (2023) *Neurosensory Divergence: Autistic Languages. A Roadmap to an Equitable Life for Autistic Children*, Authors & Co.

Feldman Barrett, Lisa (2017) *How Emotions are Made: The Secret Life of the Brain*, Houghton Mifflin Harcourt.

Ferguson, Cassie J., David A. Preece & Robert D. Schweitzer (2023) "Alexithymia in autism spectrum disorder," *Australian Psychologist*, 58:2, 131-137, DOI: 10.1080/00050067.2023.2174409

Kinnaird E, Stewart C, Tchanturia K. (2019) "Investigating alexithymia in autism: A systematic review and meta-analysis," *European Psychiatry*, Jan., 55:80-89. DOI: 10.1016/j.eurpsy.2018.09.004.

Kiraz S, Sertçelik S, Erdoğan Taycan S. (2021) "The Relationship Between Alexithymia and Impulsiveness in Adult Attention Deficit and Hyperactivity Disorder," *Türk Psikiyatri Derg*, Summer 32(2):109-117. DOI: 10.5080/u23775.

Milosavljevic B., Leno V.C., Simonoff E., Baird G., Pickles A., Jones C.R.G. (2016) "Alexithymia in adolescents with autism spectrum disorder: its relationship to internalising difficulties, sensory modulation and social cognition," *Journal of Autism and Developmental Disorders*, 46(4):1354–1367.

Milton, Damian (2012) "On the ontological status of autism: the 'double empathy problem,'" *Disability & Society*, 27:6, 883-887, DOI: 10.1080/09687599.2012.710008

Murray, D., Lesser, M & Lawson W. (2005) "Attention, monotropism and the diagnostic criteria for autism," *Autism*, May, 9(2):139-56. DOI: 10.1177/1362361305051398. PMID: 15857859.

Preece, D. A., Becerra, R., Robinson, K., Allan, A., Boyes, M., Chen, W., Hasking, P., & Gross, J. J. (2020) "What is alexithymia? Using factor analysis to establish its latent structure and relationship with fantasizing and emotional reactivity," *Journal of Personality*, 88(6), 1162–1176. <https://doi.org/10.1111/jopy.12563>

Sifneos, Peter E. (1973) "The prevalence of 'alexithymic' characteristics in psychosomatic patients," *Psychotherapy and psychosomatics*, vol. 22, no. 2-6: 255-62.

Wittgenstein, Ludwig (1980) *Remarks on the Philosophy of Psychology. Bemerkungen über die Philosophie der Psychologie*, vol. II, transl. C.G. Luckhard and Maximilian A.E. Aue, Oxford, Blackwell.

Wittgenstein, Ludwig (1992) *Last Writings on the Philosophy of Psychology. The Inner and the Outer*, vol. II, ed. G.H. von Wright and Heikki Nyman, transl. C. G. Luckhard and Maximilian A.E. Aue, Oxford, Blackwell.

Wittgenstein, Ludwig (2001) *Philosophical Investigations*, transl. G.E.M. Anscombe, 3rd edition, Oxford, Blackwell.

Organic and Artificial Intelligence: Wittgenstein's Cultural Pessimism and Modern AI

Craig Frayne (Ontario, Canada)

Abstract

What emerges from Wittgenstein's later writings is an organic, naturalistic conception of human language. Readers have contrasted this with an inorganic picture of language characteristic of the *Tractatus*. One could draw on these interpretations to critique modern artificial intelligence (AI). However, given the emergence of bio-inspired computational systems, the argument could also be made that an organic view of human languages and forms of life is not inconsistent with notions of machine intelligence. This paper discusses organic/artificial metaphors in Wittgenstein drawing from interpretations in the humanities that emphasize anti-reductionism and anti-scientism, such as that of Cavell. The paper then poses the question of whether Wittgenstein's cultural pessimism towards modern civilization is a basis for a critique of modern AI.

1. Wittgenstein's Philosophy as Cultural Metacritique

Wittgenstein's philosophy might be interpreted as a cultural metacritique of strictly logical, computational understandings of natural language. Taking language and discourse as an object of analysis is to separate it from the practices and non-linguistic forms of life into which language games are interwoven. It could be argued that digital communication exacerbates the distance between languages and forms of life, whereby natural/organic communication is usurped by symbolic imitation of human communication by machines. The notion of cultural metacritique of discourse--where language is continuously referred back to its cultural context in a form of life--could compel one could ask whether Wittgenstein's philosophy is a basis for critique of artificial intelligence.

2. Cultures as Organic Forms

Presupposing the idea of philosophy as cultural metacritique is the notion that there are cultural implications of language games and family resemblance. These implications emerge from the theme of culture in Wittgenstein's work. In certain respects, it seems obvious that culture is central to Wittgenstein's later philosophy. Following the *Tractatus* he began to see language as dynamic and learned through enculturation. His commentary on Frazer's *Golden Bough* and enthusiasm for Spengler's *Decline of the West* also indicate culture was of

great interest. However, if culture was indeed central to his thought, one could question why there seems to be little explicit philosophical attention given to the theme. It may seem reasonable to proceed by considering only those statements which do explicitly concern culture. Indeed, the aphoristic style of Wittgenstein's writing would lend itself to this selective approach. However, as Von Wright states in the preface to *Culture and Value*, such statements can be "understood and appreciated only against the background of Wittgenstein's philosophy" (Wittgenstein, 1998: ix) [Culture and Value, henceforth "CV"].

By considering the familiar philosophical concepts alongside statements in the personal notebooks, a notion of culture emerges that is both pluralistic and naturalistic. Naturalism suggests a reality independent of anyone's thoughts or worldview and may seem incompatible with the anti-realist notion of language games (Garver, 1996). Indeed, Wittgenstein emphasizes the diversity of human experiences, world pictures, and ways of using language. Underlying this diversity, however, is a shared human form of life. Arguably, our shared life-form is itself constrained by the pre-linguistic, natural world. This relation to the natural world is not one of strict correspondence or a priori understanding but is built upon layers of analogy and metaphor. In this metaphorical sense, culture stands in relation to nature as an organic form. Although Wittgenstein himself does not use this term organic form, it is here introduced to distinguish Wittgenstein's naturalism from that of the natural sciences. Organic form has a number of overlapping connotations related to artistic expression, human culture, and biological life. As a literary term, it is associated with Coleridge's idea of "unity in multeity". Goethe's morphology is also "a science of organic forms" which aims to discover "unity in the vast diversity of plants and animals" (Miller, 2009). In *Decline of the West*, Spengler (1918/65) refers to cultures through world history as the "waxing and waning of organic forms" (17–18). There is a phenomenological aspect to the concept, where understanding is obtained through first-hand experience as opposed to an external, objective point of view. It follows that grasping organic forms as wholes-- whether biological life, cultures, or works of art--is always perspectival and never ultimate. The possibility that Wittgenstein viewed languages and cultures as organic forms is a way to draw connections between his philosophy and his more enigmatic cultural views.

The idea of language as organic form can be found throughout Wittgenstein's work. This idea is expressed through metaphor, which plays a pivotal role in his later philosophy (see Gill, 1979). The *Tractatus* (Wittgenstein, 1922/2019) was concerned with forming a rather static, inorganic "picture" of language and its correspondence to the world in universal, logical space. In later writings, Wittgenstein sees language as consisting of "an inorganic part, the handling of signs" and "an organic part...understanding these signs, meaning them, interpreting them, thinking" (BB, p. 4). The *Investigations* further develops the organic notion of language: "In use it is alive" (Wittgenstein, 1986: 432) [*Philosophical Investigations*, henceforth "PI"]. It is dynamic, with parts dying off and others "coming into existence" (PI §23). Language is evolving and never complete, like an "ancient city... surrounded by a multitude of new boroughs" (PI §18). Beyond the philosophical writings, there is also indication that Wittgenstein vehemently opposed the idea of any language that had not "grown organically".

The connection between culture and organic form is based on a certain interpretation of forms of life. Forms of life can be understood as patterns and regularities "in the fabric of human existence on earth" (Pitkin, 1985). There are several interpretations of the meaning of this concept including social, cultural, behavioural, and biological accounts (see Krkač & Lukin, 2007. Sluga (2011) describes Wittgenstein's philosophy as a kind of naturalism where "forms of life, world-views, and language games are ultimately constrained by the nature of the world" (12–13). However, one could also argue that these constraints are more anthropological than biological. Keith (2012), for instance, claims Wittgenstein's position is one where there are no natural constraints on what can count truth "unless they are constraints on our shared forms of living" (487). Cavell (2013) allows for both perspectives, suggesting forms of life be seen as a relativistic "sense of agreement" as well as in a more fixed, biological sense (41–42). If forms of life are indeed constrained by the nature of the world, then there would be a unifying "system of reference" common to all human cultures (PI §206). What is common to all cultures might be activities of eating, drinking, or speaking a language (PI §25). Science and technology might also reflect transcultural, universal truths. Relativistic interpretations of forms of life counter any suggestion of universalism. However, it is important to consider that Wittgenstein does not deny the

possibility of a single or universal form of life. Sluga (2011) discusses a "single form of life" as a homogenized, unified language game and claims that, according to Wittgenstein, such a life-form would be "impoverished and almost sub-human" (61). Insofar as Wittgenstein invokes this possibility, it seems to have been a source of deep pessimism concerning the age in which he lived. Broadly speaking, this pessimism seems directed toward the scientism, positivism, and materialism he sees as characteristic of modern thought. As the antithesis of the organic diversity of language games and forms of life, Wittgenstein is a critical of the homogenizing force of modern science and technology:

Perhaps science and industry, having caused infinite misery in the process, will unite the world—I mean condense it into a single unit, though one in which peace is the last thing that will find a home. (CV: 63)

Science: enrichment & impoverishment. The one method elbows all others aside. Compared with this they all seem paltry, preliminary stages at best. (CV: 70)

The use of the word "science" for "everything that can be said without nonsense" already betrays this over-estimation. For this amounts in reality to dividing utterances into two classes: good & bad; & the danger is already there. It is similar to dividing all animals, plants & rocks into the useful & the harmful. (CV: 71)

These statements do not imply that Wittgenstein was somehow against science. Rather, they are a critique of the notion that the methods and aims of science can (or should) be applied across the range of human thought and forms of life (Read, 2016). This criticism is based on a naturalism that seeks complexity and interconnection while resisting reductionism.

3. Civilization and Cultural Decline

To further consider the idea of cultures as organic forms, we can turn to the influence of Oswald Spengler on Wittgenstein's thought. As opposed to a linear view of history and progress, Spengler depicts an organic birth and death, waxing and waning of cultures culminating in their decline as civilizations.

Civilization is the exhausted, final stage of culture. The following passage from *Decline of the West* depicts the death of culture in civilization:

Civilizations are the most external and artificial states of which a species of developed humanity is capable. They are a conclusion...death following life, rigidity following expansion, petrifying world-city following mother-earth.... The world-city means cosmopolitanism in place of "home". (Spengler: 24–25)

Cavell (1988, 2013) claims the *Philosophical Investigations* is a response to what Wittgenstein views as the Spenglerian cultural decline in the modern age. Specifically, Cavell claims that "Wittgenstein diurnalizes Spengler's vision of the destiny toward exhausted forms, toward nomadism, toward the loss of culture, or say of home, or say community" (1988: 262). According to this interpretation, Wittgenstein views the language of civilization as externalized from the language games and form of life from which it developed. Speaking outside language-games is "homologous" to the "decline of culture as a process of externalization" (Cavell, 1988: 261). In referring questions of philosophy back to ordinary language, Wittgenstein is "forgoing, rebuking, parodying philosophy's claim to privileged perspective on its culture, call it the perspective of reason (perhaps shared with science)" (Cavell 1988: 263). Explicitly, this cultural criticism is directed towards misuses of language that depart from shared forms of life. Implicitly these misuses of language are not only specific instances of decline, but homologous forms of a cultural decline that Wittgenstein's philosophy not only sought to identify, but to correct (DeAngelis, 2007). Language, specifically as it is misused or idle, is the path towards this decline. Wittgenstein confronts this situation by posing the question "is the word ever actually used in this way in the language-game which is its original home?" (PI §116). Regardless of whether the *Investigations* can be interpreted in this way, an analogous view of culture and civilization can be found in his notebooks:

It is very remarkable, that we should be inclined to think of civilization – houses, trees, cars, etc. – as separating man from his origins, from what is lofty and eternal, etc. Our civilized environment, along with its trees and

plants, strikes us then as though it were cheaply wrapped in cellophane and isolated from everything great, from God, as it were. That is a remarkable picture that intrudes on us. (CV: 50).

Perhaps one day a culture will arise out of this civilization. (CV: 74)

Others have also interpreted these statements as part of an organicist philosophy of culture. For Lurie (1989) this lament of civilization as the "taming of Nature and man" aligns Wittgenstein's thinking with the Romantic Movement (378–379). Similarly, for Pradhan (2000), Wittgenstein is expressing how "twentieth century materialist civilization" has become "detached from the springs of life and soul" (110). Cerbone (2013) claims that Wittgenstein is commenting on "something distinctively inorganic about how human beings live", analogous to his philosophy "on the *organic* and *living* character of language" (255, original emphasis). Finally, Rudd (2013) refers to Wittgenstein as a "Romantic modernist" who sought to deconstruct a way of thinking that crowds out spirit, expression, and wonder (233–234). Wittgenstein's view of culture emphasizes the diversity of worldviews and ways of using language for a given purpose. The antithesis of culture--loss of plurality and purposeful use of language--might entail a decline of an entire human form of life. The possibility that this loss was a feature of modern civilization was perhaps a source of Wittgenstein's profound cultural pessimism. The Spenglerian influence is also significant since it may imply affinity with the cultural linguistic legacies of Herder, Humboldt, Goethe, and perhaps Nietzsche. This implies that, insofar as there are opposing primordialist vs. constructionist notions of culture (see Busch, 2009), Wittgenstein could perhaps be read in the intellectual tradition of the former.

4. AI as Cultural Decline

In the contemporary global, technological context, Spenglerian notions of "world-city" and loss of "home" resonate. Perhaps a distinction between discourses of culture and those of civilization could be made. The former are expressive, symbolic, and related to dwelling in a particular time and place; the latter placeless, material, and uprooted from shared meaning. Accordingly, in international and even inter-linguistic contexts, much of the discourse of business, media, politics, and science might be understood as civilizational

rather than cultural. Unfortunately, these discourses and their associated practices too often constitute the decline, even loss, of cultures.

AI, at least in its modern manifestation, might be understood as the epitome of human culture being uprooted and replaced by inorganic, mechanized communication. It is difficult to reconcile an organic view of human cultures with a global civilization where digital data is harvested by corporations *en masse* to train machines to imitate human communication, often for commercial gain. Moreover, the rapid acceleration of AI is taking place in the backdrop of homogenization of cultures, as evidenced by the decline of minority languages (Amano, 2014).

At the same time, however, one might argue that AI is not incompatible with an organic notions of language and culture. Consider, for instance, bio-inspired algorithms and neural-nets. By conceptualizing computation as it's own 'game' embedded human forms of life (Floyd, 2012), we view AI as a reflection or imitation of the culture in which it emerges. In other words, the critique is not directed towards the notion of AI itself, but the culture an civilization in which it is embedded.

References

- Amano, T., Sandel, B., Eager, H., Bulteau, E., Svenning, J. C., Dalsgaard, B., ...Sutherland, W. J. (2014) "Global distribution and drivers of language extinction risk", *Proceedings of the Royal Society B: Biological Sciences*, 281 , 17-19. doi:10.1098/rspb.2014.1574
- Busch, D. (2009) "The notion of culture in linguistic research", *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 10.
- Cavell, S. (1988) "Declining decline: Wittgenstein as a philosopher of culture", *Inquiry*, 31 , 253-264. doi: 10.1080/00201748808602153
- Cavell, S. (2013) "This new yet unapproachable america: Lectures after Emerson after Wittgenstein". University of Chicago Press.
- Cerbone, D. R. (2013) "Dwelling on rough ground: Heidegger, Wittgenstein, architecture", in: D. Egan, S. Reynolds, & A. J. Wendland (Eds.), 245-260. Routledge. DeAngelis, W. (2007) *Ludwig Wittgenstein-a cultural point of view: Philosophy in the darkness of this time*. Ashgate.
- Floyd, J. (2012). "Turing and Wittgenstein". Retrieved from <https://www.bu.edu/cphs/files/2013/08/Floyd.pdf>

- Garver, N. (1996) "Philosophy as grammar", in: H. D. Sluga & D. G. Stern (Eds.), 139-170. Cambridge University Press.
- Gill, J. H. (1979) "Wittgenstein and metaphor", *Philosophy and Phenomenological Research*, 40 , 272-284.
- Krkač, K., & Lukin, J. (2007) "Forms of life as forms of culture", in: Hrachovec, Pichler, & Wang (Eds.), 113-115. ALWS.
- Lurie, Y. (1989) "Wittgenstein on culture and civilization", *Inquiry - an Interdisciplinary Journal of Philosophy*, 32, 375-397. doi: 10.1080/00201748908602201
- Miller, G. L. (2009) "Introduction". in: Goethe (Ed.), xv-xxxi. MIT Press.
- Pitkin, H. F. (1985) *Wittgenstein and justice: On the significance of Ludwig Wittgenstein for social and political thought*. University of California Press.
- Pradhan, R. C. (2000) "Wittgenstein and the crisis in modern civilization", *Studies in Humanities and Social Sciences*, VII , 109-118.
- Read, R. (2016) "Wittgenstein and the illusion of 'progress': On real politics and real philosophy in a world of technocracy," *Royal Institute of Philosophy Supplements*, 78 , 265-284. doi: 10.1017/S1358246116000321
- Rudd, A. (2013) "Wittgenstein and Heidegger as romantic modernists", in: D. Egan, S. Reynolds, & A. J. Wendland (Eds.), 228-244. Routledge.
- Sluga, H. (2011) *Wittgenstein*. Wiley-Blackwell.
- Spengler, O. (1965) *The decline of the west*. Modern Library.
- Wittgenstein, L. (1978) *Culture and value* (2nd ed.; G. von Wright, H. Nymen, & P. (trans.) Winch Eds.). Blackwell Ltd.
- Wittgenstein, L. (1986) *Philosophical investigations* (Third ed.; G. E. M. (trans.) Anscombe, Ed.). Blackwell Ltd.
- Wittgenstein, L. (1998) *The blue and brown books* (Second ed.). Blackwell PublishersLtd.
- Wittgenstein, L. (2010) *Tractatus Logico-Philosophicus* (C. K. Ogden, Ed.). Harcourt, Brace Company, Inc. doi: 10.1017/CBO9781107415324.004

Are Social Constructs Fictions? Odd Terminology in Harari's *Sapiens*

Martin F. Fricke (Mexico City, Mexico)

Abstract

In his *Sapiens: A Brief History of Humankind*, Yuval Harari claims that humans are able to cooperate in large numbers because they share common beliefs in fictions or “things that do not exist at all”. Examples of these fictions are religious doctrines, nations, laws, justice and money. In my paper, I argue that Harari is right to point out the importance of social constructs, entities that depend for their existence on the beliefs of the members of a society, for cooperation. But he is wrong to characterise social constructs as fictions or imagined realities. Doing so makes it difficult to distinguish between real social constructs and social constructs that are fictitious or merely imagined. Harari tries to remedy this problem calling fictitious social constructs lies; but this suggestion unhelpfully associates mere fictions or imaginings with deceptive intent.

A central claim in Yuval Harari's first bestseller, *Sapiens: A Brief History of Humankind* (2014), is that humans are distinguished from all other animals by their ability to talk about fictions, “things that do not exist at all” (Harari 2014: 27). Harari thinks that this ability is crucial for “large-scale human cooperation” (Harari 2014: 30) of thousands of inhabitants in cities or even hundreds of millions in empires. Such cooperation among large numbers of strangers is possible if they believe in “common myths” (ibid.). What are these myths? Harari mentions belief in religious doctrines, belief in the existence of a nation, of a homeland and a flag and belief in the existence of laws, justice, human rights and money. Shared belief in the value of money, for example, creates trust necessary for strangers to trade with each other. Another example are limited liability companies. They exist only because people believe that they exist, not as physical objects. But despite being only legal fictions, they are important for the functioning of the modern economy.

Yet none of these things exist outside the stories that people invent and tell one another. There are no gods in the universe, no nations, no money, no human rights, no laws, and no justice outside the common imagination of human beings. (Harari 2014: 31)

What is striking about Harari's claims is not, it seems to me, the suggestion that the cooperation of large numbers of humans depends on certain common beliefs. Rather, what surprises is his claim that these beliefs are beliefs in things *that do not exist at all*; in other words, that they are beliefs in fictions.

While contemporary readers might agree with a certain scepticism about the existence of gods, it seems odd to deny the existence of nations, money, or laws. To say that they do not exist at all or are fictions seems mistaken or, at least, in need of clarification.

Harari provides some clarification of his claim that the entities mentioned in the previous paragraph “do not exist at all”. He seems to mean that they are not physical objects: the limited liability company Peugeot SA, for example, “can’t be pointed at; it is not a physical object. But it exists as a legal entity.” (Harari 2014: 32) Here the idea seems to be that to exist, in a full sense, is to be a physical object that can be pointed at. A physical object could be defined as one that figures in physical theories; or, more traditionally, as an object with spatial location, extension, and solidity (cf. Cassam 1997: 2f.). Either way, nations, money, or laws clearly do not fall under this category and so do not exist in this sense. Nevertheless, Harari seems to admit that they exist in a looser sense, for example as legal entities. Elsewhere, he says that humans live “in a dual reality. On the one hand, the objective reality of rivers, trees and lions; and on the other hand, the imagined reality of gods, nations and corporations.” (Harari 2014: 36) He even recognises the imagined reality to have become “evermore powerful, so that today the very survival of rivers, trees and lions depends on the grace of imagined entities such as the United States and Google.” (Ibid.) It is not clear to me that this position is entirely coherent. If a reality is only imagined (“a figment of our collective imagination” [Harari 2014: 32]) and does not exist at all, how can it be powerful in our true reality? And if this is only metaphorical talk, how is it to be translated into a description of physical reality?

Whatever answers Harari might have in mind to these questions, it seems to me a mistake to characterise the above entities as *fictions*. He says that the objects in question “are known in academic circles as ‘fictions’, ‘social constructs’, or ‘imagined realities’” (Harari 2014: 35). However, while it is true that all three terms are used in academic circles with similar meanings, it seems to me that this is a mistake. We should distinguish social constructs, on the one hand, from fictions and imagined realities, on the other. The reason is that it seems plausible to say that social constructs can be either real or

fictitious; and also that they can be either real or imagined. If we conflate social constructs with fictions and imagined realities, it becomes impossible to make these distinctions with respect to social constructs.

According to some philosophers, elements of social reality exist because and in so far as people believe or accept that they exist. They depend for their existence on people's beliefs (Searle 1995). For instance, a judge is a judge because people accept the laws that define the requirements for becoming a judge and they accept the judge's rulings as valid and only revisable through certain procedures. These elements of social reality are often called social constructs. Thus, social constructs, and social reality in general, are belief-dependent. They do not exist independently of people's beliefs in the way physical objects exist.

Does this mean that social constructs are fictions? It seems to me that it does not; for social constructs can either be real or fictitious. Let's take Harari's example of the limited liability company. Before the legislation was passed which makes it possible to create such companies and before someone took advantage of this legislation and created such a company, they did not exist and were not real. However, some intelligent person might nevertheless have written a fictional account of a society in which it was possible to create a limited liability company and in which people took advantage of this possibility and did create such companies. These would have been fictional limited liability companies, since they would not have existed in the real world, only in fiction. Both a fictional and a real limited liability company are social constructs. They exist only because people believe that they exist. But it seems plausible to say that only one of them is a fiction and only one of them is real. If we insist on calling *all* such companies fictions, we cannot draw this distinction.

Likewise, we might imagine a society in which it is possible to enter a marriage of up to three people of the same or different sexes. Such a marriage would be a social construct, but it would only be an imagined reality, not part of the real world. In the real world, marriage is also a social construct, but – at least in Kirchberg – it involves only two people. The difference between the two cases is that the first type of marriage is an imagined social construct or an imagined reality, while the second is a real social construct. If we characterise *all*

marriages as imagined realities, it becomes impossible to distinguish between the two cases.

So while it might make sense to call at least some of the constituents of the social world social constructs, because unlike physical objects, they are dependent on people's beliefs, it seems to me that we should not call them fictions or imagined realities, because this would make it impossible to distinguish between the real social world and a fictitious or an imagined one.

Harari tries to solve this problem with his terminology by distinguishing between lying and sincere belief: If we talk about what I called a fictitious or an imagined social construct, we are lying, according to Harari. By contrast, talk about what I called real social constructs is based on sincere belief.

An imagined reality is not a lie. [...] Unlike lying, an imagined reality is something that everyone believes in, and as long as this communal belief persists, the imagined reality exerts force in the world. [...] Some sorcerers are charlatans, but most sincerely believe in the existence of gods and demons. Most millionaires sincerely believe in the existence of money and limited liability companies. (Harari 2014: 35)

Is this a reasonable solution to the terminological problem? It is a solution of sorts because at least it allows to distinguish between social constructs that are real and those that are not. Both might be imagined realities or fictions, in Harari's terminology; but the second ones are also lies. However, this is a solution with flaws. First, to say that *all* social constructs are fictions or figments of our imagination suggests that our real social world somehow lacks reality, that it "does not exist at all". This might be a spectacular claim useful for writing a bestseller, but it is also disingenuous if all we want to say is that social constructs depend for their existence on certain shared beliefs and acceptances in a society. Belief-dependent objects are no less real than physical objects. They merely have different persistence conditions.

The second flaw is that this terminology associates what we ordinary call fiction or imaginings with lying. As Harari himself explains, someone who lies lacks sincerity. She tries to *deceive* her listeners. But this is an inadequate characterisation of those who – as we ordinarily would say – write fiction. García Marquez writes about things that do not exist in his novels. But his isn't

to say that he is trying to deceive us. He simply writes fiction, rather than trying to give an objective description of reality. Likewise, it is often a useful exercise to imagine a world that does not (yet) exist. We can try to imagine what it would be like if everyone enjoyed a universal basic income. Would this be a good thing or a bad thing? Such counterfactual reasoning can be seen as a way of discovering what is possible in the real world (Williamson 2007). Yet it seems that Harari would have to say that it is based on lies, on intents of deception. This would seem an odd characterisation of such reasoning, especially given that the only reason he has for making it is his desire to declare that social constructs are merely imagined and somehow lack reality.

To sum up. Harari make the plausible claim that large numbers of humans can only cooperate if they share common beliefs, which help to constitute their social reality. He makes the further claim that the social constructs which depend on these common beliefs are merely fictions or imagined realities and as such “do not exist at all”. Although he qualifies the “not existing at all” claim somewhat, it is at least terminologically awkward to characterise social constructs as fictions or imagined realities, because it makes it difficult to distinguish between social constructs that are real in our social realities and social constructs that are fictitious or merely imagined, such as in counterfactual reasoning. Harari tries to remedy the problem by distinguishing between sincerely believed social constructs and lies about them. But this only creates further problems, as it suggests that social reality does not exist at all and associates what we ordinarily know as fiction and imagining with deceptive lying. We can avoid all these problems by not taking social constructs to be mere fictions or imagined realities.

References

Cassam, Quassim (1997) *Self and World*, Oxford: Clarendon Press.

Harari, Yuval Noah (2014) *Sapiens. A Brief History of Humankind*, London: Penguin Random House.

Searle, John (1995) *The Construction of Social Reality*, London: Penguin Random House.

Williamson, Timothy (2007) *The Philosophy of Philosophy*, Malden, M.A.: Blackwell.

On Raising My Arm Automatically and Similar Actions: Implications of Automaticity for Free Agency

Christina Fritz (Graz, Austria)

Abstract

This paper provides novel perspectives on human agency. In particular, I aim to direct focus on the fact that most of the actions we perform in our daily lives are performed automatically. That is, we perform these actions without (constantly) paying attention to them. Importantly, in automatic actions, we are unaware of any decision-making process or deliberation preceding the action. This raises the question of whether automatic actions are governed by decisions and reasons in the same way as deliberate actions. As it is tempting to think of deliberation as being necessary for freedom of action, the nature of automatic actions has crucial implications. In particular, if (i) only actions based on deliberation can be free and (ii) most of our actions are automatic actions, this seems to imply that we act unfreely most of the time. By critically questioning Carolina Sartorio's promising account of an actual-sequence view of freedom, I would like to make it clear that automatic actions are also rooted in the decisions of the agent, albeit these decisions occur at a subconscious level.

1. Introduction

Oops! ... I Did It Again (Spears 2000). – This is how to grasp automatic actions in a nutshell. Just as you can't help but have a classic pop song stuck in your head at this moment (you're welcome!), you can't help but act automatically. We perform manifold automatic actions every day. Examples of these specific types of actions range from scratching our heads inattentively to more complex actions such as those involved in driving a car. We cannot completely avoid acting automatically. We can interrupt and control automatic actions to a certain extent, but not indefinitely.

Automatic actions seem to be little studied in philosophical research. Mainstream theories of the philosophy of action tend to focus on deliberate actions of which we are fully aware, while neglecting the bulk of actions that we tend to perform on autopilot. Consequently, we generally tend to think about action in the following way (roughly speaking):

Person *x* thinks about what to do next.

Person *x* comes to a decision.

Person *x* acts on the basis of her decision.

This is deliberation. When a person deliberates and acts on the basis of her deliberation, we would say that the person acts out of free will. Consequently,

it is tempting to think of deliberation as being necessary for freedom of action. When we consider the nature of automatic actions, we must soon admit that maintaining deliberation as a necessary condition for free agency has serious consequences.

The main problem arises as follows: On the one hand, deliberation is defined as a process of thoughtfully and consciously weighing upon reasons, while on the other hand automatic actions are typically characterized in the opposite way. For instance, in his influential paper *Mental Action and the Threat of Automaticity* Wayne Wu offers the following differentiation: “Automatic processes are contrasted with deliberate, attention-demanding, conscious, controlled aspects of cognition” (Wu 2013: 246). Assuming such a distinction between deliberate and automatic actions, this leads to the following challenge: If (i) only actions based on deliberation can be free and (ii) most of our actions are automatic actions, this seems to imply that we act unfreely most of the time. This challenge is particularly pressing since scientific research shows that actually “only 5% of our actions are consciously controlled, the rest being automatic” (Lumer 2017: 3; see, e.g., also: Bargh; Chartrand 1999, Custers; Aarts 2010, Schlosser 2019).

In the following, I draw on the vivid example of raising one’s arm to illustrate the distinction between raising it deliberately and raising it automatically. In this respect, Ludwig Wittgenstein provides us with a striking example that will guide us through my analysis. My objective is to illuminate the significant implications of raising my arm automatically and similar actions. Consequently, I will argue that even actions performed automatically retain a share of prior deliberation and thus can be regarded as actions that were performed freely.

2. Wittgenstein’s puzzle extended

In theories within the philosophy of action and regarding freedom of will, when deliberation is mentioned, it is assumed that the mental antecedent of an agent’s action (the act of deliberation) is conscious. Consequently, some may not acknowledge automatic actions as genuine actions, as they occur

subconsciously and without the agent's awareness of a preceding act of deliberation. I argue that automatic actions are indeed full-fledged actions caused by mental antecedents that are themselves subconscious.

Consider a classical example famously made by Wittgenstein: “[W]hat is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?” (PI 1953: §621). Clearly, the act of raising the arm belongs to the agent who is performing it. This action is caused by the agent, irrespective of whether it occurs with full consciousness or automatically. However, is *what is left over* the same with automatic actions as opposed to consciously performed ones? There is certainly disagreement.

A rough distinction between “action” and “automatic action” could be the following:

ACTION: An event brought about by an agent as a result of a specific mental cause (i.e., an act of deliberation).

Concerning automatic actions there is no doubt that they were also brought about by agents. However, there is severe disagreement regarding the nature and even existence of the corresponding mental causes. Three options are specified below:

AUTOMATIC ACTION: An event brought about by an agent, the corresponding mental cause being (i) subconscious, and/or (ii) derived from prior conscious appearances of mental causes, or (iii) not present at all.

Note that option (iii) carries significant implications as it contradicts the definition of action. In this context, automatic actions would not be considered actions at all. However, options (i) and (ii) are interconnected, and these are the options I advocate for. It is only through options (i) and (ii) that we can comprehend automatic actions as genuine actions according to the provided definition of action.

Before I present my proposed solution, which argues that automatic actions should be considered genuine actions according to the definition provided above, the following two subsections aim to explore the analysis of raising one's arm in greater depth. Here I will highlight the most significant

implications as I perceive them, focusing on both ethical and epistemological considerations.

(a) Ethical implications of raising my arm automatically

Strongly connected with questions about free will of an agent are questions about her moral responsibility. The commonsense view is that an agent is morally responsible for her actions if they are the outcome of her own decisions rather than coercion and the like. It seems essential to take this seriously. In our society and within legal affairs it is clearly relevant whether a person did what she did on the basis of her own decisions and plans or not. The view that not only moral responsibility but also free will depends on having our actions grounded in the right causal histories, i.e., decisions and reasons of the agent, is called the actual-sequence view of freedom. I'll come back to that in greater detail in the last section of this paper.

Let's now consider an automatic action that, initially, carries no moral weight, such as raising my arm to scratch the back of my head. If this action happens completely inattentively, leading to accidentally hitting someone's face, it's reasonable that the person who was hit will blame me for getting hurt. Most would agree that such a reaction is justified. While the level of responsibility might be lower compared to deliberately hitting someone, it doesn't absolve one of responsibility entirely. A common defense might be, "I'm sorry, I was lost in thought!"—but the response could easily be, "Then pay attention!" because we could have watched out and paid attention as we could have changed our behavior. The automatic raising of one's arm is still attributed to the agent; the arm did not simply *go up*. Therefore, it's understandable to blame someone for her actions even in such circumstances in which the agent certainly had acted differently if she had paid attention. Also, as we can see, automatic actions are particularly challenging when it comes to an assessment of the moral responsibility of an agent. These actions clearly belong to us but the amount of control we have over these actions is much less than when we deliberate. Let us keep these implications for moral responsibility in mind while looking at the epistemic implications of acting automatically.

(b) Epistemic implications of raising my arm automatically

Now I address the veridicality of agentive experiences in the context of automatic actions. Imagine that you are experiencing yourself raising your

arm. You can observe your arm rising as you feel the movement. You have your reasons for raising your arm, perhaps to ask a question. Nothing out of the ordinary has occurred—no witches, no sinister scientists, no god-like beings causing illusions. The sensation of your arm moving is no different from any other experience of moving your body. There's no reason to doubt your senses.

Would you consider your experience in this case to be veridical? In other words, are you justified in believing that you are raising your arm? It is widely accepted that you are justified in believing that your arm is indeed raising and that you are the one causing it to do so. Consequently, most of us would agree that experiences of this kind are veridical rather than illusory.

However, when an agent is acting in an automatic manner, she is not experiencing herself acting in the same way. Still, these are actions. The raising of the arm does belong to the agent who is raising her arm, it does not *go up* independently of the agent. Terry Horgan and Martine Nida-Rümelin argue that agentic experiences have a distinctive phenomenology of agency and that these experiences are typically veridical as well as justifying. They only discuss this in the context of deliberate actions though (see Horgan; Nida-Rümelin 2021).

As automatic actions are clearly *ours* and we can pay attention to them whenever necessary, they are part of the overall phenomenology of agentic experiences, they are part of experiencing ourselves as agents. Here is a convincing example concerning the phenomenology of visual experiences which helps to see what I have in mind concerning experiencing automatic actions:

Consider, for example, your own visual experience. Looking straight ahead, can you see your nose? It might initially be tempting to answer in the negative, or by saying that you are not sure (after all, the 'edge' of the visual field is a surprisingly difficult thing to describe). But now wiggle your nose. I suspect that something moves within your field of vision. Or close one eye. Now do you see it? If you are anything like me (and I don't think that my nose is unusually long) you can see your nose, but it is presented in what one might describe as a reclusive way. (Smith 2016: 5)

I believe that similarly to how having a nose inattentively shapes the phenomenology of our visual experiences, performing automatic actions inattentively shapes the overall phenomenology of our experiential lives. Since we perform automatic actions all the time, but typically do not experience ourselves as being the agents of these actions, this reveals challenges that need to be dealt with in order to better understand ourselves, how we act in the world, and how we affect our fellow human beings. For reasons of space, I cannot do justice to this here and will now proceed to my conclusion by refining the actual-sequence view of freedom in the light of automaticity in action.

3. A refinement for the actual-sequence view of freedom

If we do not want to accept that most of our actions lack genuine agency and thus cannot be free, we need to aim at a more fine-grained analysis of agency that allows considering automaticity and freedom to be consistent. To this end, I focus on Carolina Sartorio's compatibilist approach of an *actual-sequence view of freedom*. I pick Sartorio's approach primarily because Sartorio explicitly emphasizes to understand "freedom to be only a function of the actual sources or the actual causes of action" (Sartorio 2017). I find her view highly plausible, but it seems to be inconsistent with automatic actions being performed freely. I will show how her approach can be modified such that freedom and automaticity are made compatible.

The actual-sequence view of freedom offers a novel response to the question of whether determinism and freedom are compatible. First advocated by John Martin Fischer and further developed by Mark Ravizza, the actual-sequence view of freedom claims that freedom of action is a matter of having actions grounded in the right causal histories. It can be contrasted with the principle of alternative possibilities which claims that freedom of action is a matter of having actions grounded in the possibility to act otherwise. This approach was famously challenged by Harry Frankfurt. The well-known Frankfurt cases can be seen as fundamental for the actual-sequence view of freedom. Here is what a common Frankfurt case scenario looks like:

Frank and Furt: Frank is deliberating about whether to shoot Furt. Unbeknownst to him, a neuroscientist has been secretly monitoring

Frank's brain processes. The neuroscientist can reliably predict the choices that Frank is about to make by looking at the activity in his brain, and can also manipulate Frank's brain in a way that guarantees that Frank will shoot Furt. He plans to intervene if he predicts that Frank will not choose to shoot Furt on his own. As it happens, Frank chooses to shoot Furt on his own, motivated by his own reasons, and without the intervention of the neuroscientist (who correctly predicts that Frank would make that choice on his own). (Formulated after Frankfurt 1969 by Sartorio 2016 and Kaiserman 2020: 2).

We see, in this scenario the subject had no real alternative possibility in his acting: he wasn't able to do otherwise, because the slightest flicker of thought that could possibly have changed his mind would have provoked the neuroscientist to manipulate Frank's brain. However, Frank did decide on his own to shoot Furt: "motivated by his own reasons". The principle of alternative possibilities fails when applied to the Frankfurt cases. Note that the Frankfurt case is applied to the freedom of an agent and not only her moral responsibility. It clearly shows that Frank did decide on his own and that his action was motivated by his own reasons. His action was grounded in the right causal histories.

Sartorio reacts directly to Frankfurt cases and brings *deliberation* as part of the right causal histories back into play:

In a Frankfurt case, an agent makes a choice completely on his own, based on his antecedent deliberation and careful weighing of reasons, and thus the agent appears to have the relevant kind of freedom or control to be morally responsible for it. (Sartorio 2015: 105)

Now, it's essential to gain clarity regarding the right causal histories of free action. Is there a decision preceding every single (and also automatic) action? Or is the decision of an agent causing already a whole bundle of actions? I am suggesting that every individual action has a share of an earlier decision-making process or deliberation that has set in motion a whole train of actions in the first place. Automatic actions, as I am referring to them, are actions performed automatically and without being (fully) aware of acting while performing these action(s). The dimension of not being aware of an automatic

action opens up at least two or perhaps three different issues that we should keep apart:

1. In automatic actions, the agent is not consciously aware of performing the action itself.
2. In automatic actions, the agent is not aware of the decision-making process that led to the action.
3. In automatic actions, the agent is not aware of the influences of previous decisions on their current actions.

While we quickly recognize (1) when considering automatic actions, it is crucial to consider (2) and (3) in more detail. Questions about (2) might lead us to answers about whether an agent acts freely when performing automatic actions. Questions about (3) also reveal consequences for the moral responsibility of an agent in the context of automatic actions.

Here I focus on (2). I aim to make clear that automatic actions are based on decisions, but these decisions are subconscious, often involving subconscious reasoning. We are making decisions (reasoning occurs), but we are not aware of them. Consequently, I suggest a differentiation of *what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm*; I suggest a differentiation of the will: There is the *occurrent* decision at the moment one consciously chooses to act: choosing to sit down and get some work done and the like. This is in contrast to the many *latent* wantings that occur subconsciously but as a consequence of our previous *occurrent* decisions.

My thesis is that automatic actions are based on subconscious decision-making processes which can be illustrated by giving examples of latent wantings, such as “wanting to stay alive”, “wanting to stay safe”, “wanting to feel good” – several wantings that we are not aware of at every moment, but which are manifoldly present. Basically, such a differentiation of our will in terms of an *occurrent* decision in contrast to *latent* wantings is crucial to successfully implement automatic actions within a comprehensive actual-sequence view of freedom. In this picture, even automatic actions can be grounded in the right causal histories and thus can be considered to have been performed freely. Although in the case of automatic actions the agent does not have an “antecedent deliberation and careful weighing of reasons” right before each

individual action, the freedom of the individual automatic action depends on an earlier decision-making process. If the previous decision-making process (including the occurrent decision), which triggered the entire sequence of actions, was free, then the subsequent (automatic) actions are also free. I think this offers a small but important piece of the puzzle for allowing free will (and moral responsibility) to be compatible with automaticity in action.

References

- Bargh, John A.; Chartrand, Tanya L. (1999): "The Unbearable Automaticity of Being", *American Psychologist*, 54 (7), 462-479.
- Custers, Ruud; Aarts, Henk (2010) "The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness", *Science*, 329 (5987), 47-50.
- Frankfurt, Harry G. (1969) "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, 66 (23), 829-839.
- Horgan, Terry; Nida-Rümelin, Martine (2021) "On the Satisfaction Conditions of Agentic Phenomenology: a Dialogue", in: Erhard, Christopher; Keiling, Tobias (eds.): *The Routledge Handbook of Phenomenology of Agency*. Routledge.
- Kaiserman, Alex (2020) "Reasons-sensitivity and Degrees of Free Will", *Philosophy and Phenomenological Research*, 103 (3), 687-709.
- Lumer, Christoph (2017) "Automatic actions: Agency, Intentionality, and Responsibility", *Philosophical Psychology*, 30 (5), 616-644.
- Sartorio, Carolina (2015): "Sensitivity to Reasons and Actual Sequences", in: Shoemaker, David (ed.): *Oxford Studies in Agency and Responsibility: Volume 3*. Oxford/New York: OUP.
- Sartorio, Carolina (2016) *Causation and Free Will*. OUP.
- Sartorio, Carolina (2017) "Actual Causes and Free Will", *Disputatio*, 9 (45): 147-165.
- Schlosser, Markus (2019) "Agency", in: Zalta, Edward N. (ed.): *The Stanford Encyclopedia of Philosophy*. Winter 2019 Edition. URL: <<https://plato.stanford.edu/archives/win2019/entries/agency/>>.
- Smith, Joel (2016) *Experiencing Phenomenology: An Introduction*. Routledge & CRC Press.
- Spears, Britney (2000) *Oops!... I Did It Again*. In her second studio album of the same name.
- Wittgenstein, Ludwig (1953) *Philosophical Investigations / Philosophische Untersuchungen*. Anscombe, G.E.M.; Rhees, R. (eds.), translated by Anscombe, G.E.M., Oxford: Basil Blackwell.

Wu, Wayne (2013) "Mental Action and the Threat of Automaticity", in: Clark, Andy; Kiverstein, Julian; Vierkant, Tillmann (eds.): *Decomposing the Will*. OUP.

Acknowledgement

For their many helpful suggestions, I am incredibly grateful to Katharina Felka, Philipp Berghofer and Marian David from the University of Graz.

The Regress of Necessity and Its Structure

Yaokun Fu (Vienna, Austria)

Abstract

In this paper, I argue that those relying solely on the transmissive model to explain necessity face a regress of necessity. I then argue that the regress of necessity is vicious due to explanatory failure because it fails to address the source question of necessity adequately. We have prima facie justification for this conclusion in the absence of positive arguments for necessity infinitism, which holds that an infinite chain of necessity could serve as the source of necessity. On pain of vicious regress, we should accept necessity foundationalism, which posits the existence of foundational box-truths (truths of the form $\Box p$) whose necessity is not explained by the necessity of further box-truths. Necessity foundationalism is motivated by the regress of necessity in a similar manner to justification foundationalism, which is motivated by the regress of epistemic reason.

1. Introduction

Bob Hale (2002, 2013) once motivated the existence of foundational necessary truths on the ground that a non-transmissive explanation of necessity is possible. The non-transmissive explanation explains the necessity of a necessary truth by appealing only to the truth of another necessary truth without its necessity doing any explanatory work. In contrast, the transmissive explanation appeals to the necessity of further necessary truths. However, Carlos Romero (2019, forthcoming) doubts that the non-transmissive explanation is a genuine possibility. In this paper, I shall provide a novel argument for the existence of foundational necessary truths, which, in my terminology, amounts to accepting necessity foundationalism.

Here is some jargon for further discussion. Truths that take the form " $\Box p$ " are called *box-truths*. Further box-truths are appealed to in a transmissive explanation of a box-truth, and the former box-truth is explained transmissively. In a non-transmissive explanation of a box-truth, no appeal is made to further box-truths, and the box-truth is explained non-transmissively. *The transmissive model* explains every box-truth transmissively, while *the non-transmissive model* explains some box-truths non-transmissively.

Here is my plan. Section 2 begins with Simon Blackburn's (1993) dilemma for explaining necessity and demonstrates how there is a regress of necessity for those who believe the only plausible model to explain necessity is the transmissive one. Section 3 proceeds to argue that the regress of necessity is

vicious by virtue of explanatory failure, applying Michael Huemer's (2016) theory of viciousness. In Section 4, I motivate necessity foundationalism on pain of vicious regress.

2. The regress of necessity

Simon Blackburn (1993: 53) introduces a dilemma in explaining necessity. Suppose we attempt to explain the box-truth $\Box p$. We have two options: either we appeal to something contingent (the contingency horn) or something necessary (the necessity horn). The core idea of the contingency horn is that contingent explanans are not suitable candidates for explaining box-truths. For a detailed discussion of the contingency horn, refer to Hale (2002) and Hale (2013), chap. 3.5. (I shall not go into details here.) If the explanans is necessary, it fails to provide the desired explanation because it merely shifts the question to another box-truth.

Bob Hale (2002) noticed that the necessity horn rests on the transmissive model, which invokes a further box-truth to explain every box-truth.

A transmissive explanation has the form: necessarily B because [necessarily A and it follows from A that B]—the necessity of A is transmitted across the entailment to the explanandum. (Hale 2013: 96. My brackets for clarification of scope.)

He contended that the transmissive model is not the only possible and proposed that the non-transmissive model is also possible. According to the non-transmissive model, some box-truths are explained non-transmissively.

A non-transmissive explanation is an explanation of the form ' $\Box p$ because q ' in which the explanans, q , is indeed necessary (at least if it is true, as it must be if we are to have an explanation at all), but in which what explains the necessity of the explanandum is not q 's necessity, but its truth simpliciter. (Hale 2013: 131)

However, some philosophers, like Romero (2019), might doubt that the non-transmissive is possible. Romero (forthcoming: 10) emphasizes, "I do not think it is obvious that necessity can play no role in the explanation—one may doubt that there truly are non-transmissive explanations." As he speculated in his

earlier paper (2019: 127), "... explanations of modality seem to be non-transmissive only because the modal principle of the necessity of [the explanans] ...is implicitly assumed." (My brackets for simplifying discussion) If Romero argues against the possibility of the non-transmissive explanation, it is plausible to interpret "the modal principle" as "the explanandum's necessity somehow follows from the explanans's necessity". Then, the non-transmissive explanation is impossible because every non-transmissive explanation is implicitly transmissive in the sense of assuming "the modal principle". Therefore, it seems that the non-transmissive model is not plausible.

However, an infinite regress of necessity would ensue if the only plausible model for explaining necessity is the transmissive one. Once again, suppose the box-truth $\Box p$ needs explanation, and only the transmissive model is available. We appeal to another box-truth $\Box q$. Since the necessity of $\Box q$ plays an explanatory role, the necessity of $\Box q$ is called into question. We then introduce another box-truth $\Box r$ to explain the box-truth $\Box q$, leading to the necessity of $\Box r$ needing explanation. Again, we can only appeal to another box-truth, and an infinite regress of necessity would ensue if there were no principled way to stop the reasoning.

One might question why the regress of necessity is considered a serious problem if we are unsure whether the infinite regress is vicious or benign. In the following discussion, I will evaluate the regress of necessity and argue that it is indeed a vicious one.

3. The regress of necessity is vicious

To diagnose the viciousness of a regress, it is desirable to identify common features shared by most vicious regresses. I will apply Michael Huemer's (2016) theory of viciousness to diagnose the regress of necessity. Huemer (2016: 229–246) identifies three features of viciousness: metaphysical impossibility, extreme implausibility, and explanatory failure (Ricki Leigh Bliss (2013) also considers explanatory failure as an important feature of vicious regressions). Only the last one is relevant to our discussion. This is because, firstly, an infinite regress is considered vicious due to a metaphysical impossibility, according to Huemer's account, only if it implicates an infinite natural intensive magnitude (such as mass or energy). Presumably, the regress

of necessity does not involve an infinite intensive magnitude, so this feature of viciousness is irrelevant here. Secondly, an infinite regress is deemed vicious by being extremely implausible, mainly when it is empirically implausible that humans can carry out such an infinity. This feature primarily relates to regressions involving human capacities. For example, in the regress of epistemic reason, it is empirically implausible that humans can carry out such an infinite series of reasons to justify any beliefs.

The Regress Argument for (Justification) Foundationalism: "This argument claims that because there cannot be an infinite series of reasons for any belief, there must be some beliefs that are justified in a way that does not depend on reasons. These 'foundational' beliefs would be the source of the justification for all other justified beliefs." (Huemer 2016: 229. My brackets for clarification.)

The notion of explanation featured in the regress of necessity does not concern what human beings could do (it is not an epistemic notion but a metaphysical one). Therefore, extreme implausibility is irrelevant to the regress of necessity.

What remains, then, for an infinite regress to be considered vicious is its indication of a theory's failure to explain what it intends to explain. Let me begin with two simple (but false) scientific theories and then, by examining them, derive a general procedure to determine when an infinite regress is vicious due to explanatory failure.

These two examples (from Huemer (2016) and Bliss (2013)), which characterize explanatory failure, are the homuncular theory of perception and the turtle's regress. First, let's consider the homuncular theory of perception: According to this theory, for a person to see an object just is for a tiny person sitting behind the person's eyes to receive and process information and send it to the brain. However, this leads to the question: why does this tiny person have vision? The theory posits that a second little person sits behind the first one's eyes to receive and process what the first one sees. This cycle continues indefinitely, resulting in an infinite regress of little persons sitting behind eyes.

This theory is false but a good example of explanatory failure. It aims to explain how vision happens by positing a tiny person behind the eyes. However, this approach raises questions about the tiny person's vision, which

also needs explanation. Notably, the problem with the homuncular theory isn't solely about positing an infinite number of tiny persons. Even if we entertain the idea of an infinite number of tiny persons inside our brains, the theory still falls short of explaining how vision happens due to explanatory failure at every level. At each level of explanation, a tiny person's vision remains unexplained. While the vision of each particular tiny person may be successfully explained, dissatisfaction arises when considering the general vision process. This dissatisfaction stems from the consistent reappearance of vision in the explanans.

The second example is the turtle's regress. Consider the question, "How could the earth not fall down in space?" A pre-scientific cosmological theory says, "There is a giant turtle supporting the earth." We wonder what happens to the giant turtle and why it does not fall. The response is, "It's turtles all the way down."

The turtle's regress is considered vicious because it fails to explain why the earth does not fall down in space. By positing the first turtle, the state of not falling reappeared in the explanans, and the first turtle's state of not falling requires further explanation. It is not helpful to explain why the earth does not fall even by introducing an infinite number of turtles because, at every level of explanation, there is always a turtle's state of not falling that remains in question and unexplained.

At first glance, one might think an infinite series of turtles could explain why the earth does not fall down in space. However, as Huemer (2016: 237) suggests, it could be an infinite series of falling turtles rather than an infinite series of stationary turtles. Thus, even an infinite chain of turtles as a whole still falls short of explanation.

Now, I am in a position to propose a general procedure for determining when an infinite regress is vicious due to explanatory failure:

1. If there is an infinite regress, identify whether there is a valid explanatory task.
2. If there is a valid explanatory task, identify whether an infinite regress could fulfil the explanatory task. To appreciate how this might be done, consider justification infinitism, which claims that the source of

justification is an infinite, non-repeating chain of epistemic reasons standing behind each justified belief.

What about the regress of necessity? My answer: it is vicious by virtue of explanatory failure. The first step in our analysis is identifying the explanatory task. Rather than explaining particular box-truths, the crux lies in explaining necessity in general. Hale (2002: 309) sometimes takes an explanation of necessity in general as addressing an "anything at all" question, namely, "Why is there necessity at all?" which amounts to explaining why it is true that there is at least one necessary truth. If we interpret the explanatory task this way, it will have only simple answers. Following most of the literature on grounding, I take existential truths to be grounded in their true instances. Then the question "Why is there necessity at all?" has some simple answers: because it is a necessary truth that Donald Trump is self-identical, and because it is a necessary truth that I cannot have been born of different parents. (Thank Julio De Rizzo for pressing this line of reasoning to me.)

If these simple answers prove unsatisfactory, we must reconsider our interpretation. Another plausible interpretation of the "anything at all" question is to take it as a query about the source of necessity. I do not delve into a detailed exploration of the source question because I trust my readers to find it intelligible. Prominent answers to the source of necessity in literature include our linguistic conventions, essences of things, and so on. In epistemology, the source question of justification is already familiar. A notable answer from foundationalists is appearance or seeming (Huemer 2007).

The second step is to determine whether the regress of necessity could fulfil the explanatory task of answering the source question. It falls short because there is always an unexplained necessity at every level of explanation, and the residual necessity reappears infinitely many times within the explanans. Thus, the regress of necessity cannot answer the source question.

It is plausible to conclude that the regress of necessity is vicious due to explanatory failure. The vicious regress poses a severe problem to those who believe that the only plausible model for explaining box-truths is the transmissive one, as its viciousness indicates the impotent explanatory power of this model. The best approach to avoid the regress of necessity is to accept

the non-transmissive model as a genuine possibility. As I shall argue next, the regress of necessity independently motivates what I will label "necessity foundationalism".

4. Necessity Foundationalism

Necessity foundationalism is the thesis that there are foundational box-truths whose necessity (i) is not explained by the necessity of further box-truths, (ii) can explain the necessity of every box-truths distinct from them, and (iii) is non-transmissively explained by further necessary truths. I admit that the definition is informal. But if the above "explain" could be cashed out in terms of grounding, a formal definition is in the vicinity. However, I leave that task for another paper.

Why accept necessity foundationalism in the first place? It is because it offers a way to avoid the regress of necessity, and the vicious nature of the regress serves as motivation for it. Thus, it seems that a vicious regress could indeed serve as an argument for a specific form of foundationalism. This is evident in justification foundationalism, which posits the existence of basic beliefs—sometimes called foundational beliefs—that are justified in a way that does not depend on being justified by other beliefs.

As mentioned earlier, the regress of epistemic reason is vicious by virtue of extreme implausibility, and it serves as motivation for justification foundationalism. An analogy can be drawn between the regress of epistemic reason and the regress of necessity because they are both considered vicious due to explanatory failure. (This suggests that a regress could be vicious in two distinct senses.)

The regress of epistemic reason is supposed to answer the source question of justification. Recall justification infinitism, which holds that for a belief to be justified is for it to have an infinite, non-repeating chain of reasons standing behind each justified belief (Huemer 2022: 81). However, justification infinitism fails to answer the source question because at every level of reasoning, there is always an unjustified belief reappearing within the explanans. Even if an ideal agent could reason through an infinite series of chains, the regress of reason as a whole still falls short of answering the source

question of justification. Therefore, the regress of epistemic reason is vicious by virtue of explanatory failure, rendering justification infinitism unmotivated.

Let us refer to the theory that accepts the regress of necessity (as benign) as necessity infinitism. The theory proposes that for every box-truths, there are infinitely many box-truths behind it, serving as the source of necessity. From our previous discussion, without plausible arguments for necessity infinitism, it is at least *prima facie* justified for us to conclude that the regress of necessity is vicious due to explanatory failure and necessity infinitism is unmotivated. The plausible theories left for us are justification foundationalism and necessity foundationalism. Therefore, just as the regress of epistemic reason could motivate justification foundationalism, the regress of necessity could motivate necessity foundationalism. (It's worth noting that a complete analogy would involve mentioning justification coherentism and necessity coherentism as well—maybe even foundherentism—but due to our limited discussion space, I'll reserve that for another context.)

References

- Blackburn, Simon (1993) "Morals and Modals", in: *Essays in Quasi-Realis*, New York: Oxford University Press, 52–74.
- Bliss, Ricki Leigh (2013) "Viciousness and the Structure of Reality", *Philosophical Studies* 166 (2), 399–418.
- Hale, Bob (2002) "The Source of Necessity", *Noûs* 36 (Suppl 16), 299–319.
- Hale, Bob (2013) *Necessary Beings: An Essay on Ontology, Modality, and the Relations Between Them*, New York: Oxford University Press.
- Huemer, Michael (2007) "Compassionate phenomenal conservatism", *Philosophy and Phenomenological Research* 74 (1), 30–55.
- Huemer, Michael (2016) *Approaching Infinity*, New York: Palgrave Macmillan.
- Huemer, Michael (2022) *Understanding Knowledge*.
- Romero, Carlos (2019) "Modality is not Explainable by Essence", *Philosophical Quarterly* 69 (274), 121–141.
- Romero, Carlos (forthcoming) "On Explaining Necessity by the Essence of Essence", *Inquiry: An Interdisciplinary Journal of Philosophy*.

Acknowledgement

I would like to thank Benjamin Schnieder and Julio De Rizzo for their comments and discussion.

Two Concepts of Possibility in *Tractatus Logico-Philosophicus*

Cloris C. Gao (Iowa, USA)

Abstract

Can we imagine that some basic objects in a state of affairs are to be removed from that situation? In *Tractatus Logico-Philosophicus*, Wittgenstein answers both “yes” and “no”. According to Peach (2007), this tension is created by the ambiguity of the two concepts of possibility in *Tractatus*: the object-based concept of possibility, which is contained in a simple object’s possible occurrences in states of affairs, and the “space metaphor” notion of possibility, which is revealed in the imagination of an empty logical space. This paper examines these two concepts of possibility and their relationships to the two concepts of space, namely, the Newtonian idea of space and the Leibnizian notion of space. According to Newton, the idea of a space is independent from the objects it contains and thus can be empty, whereas for Leibniz, the concept of space cannot be separated from the individual substances. In this paper, it is suggested that Wittgenstein may also embrace Leibniz’s idea of space, given the similarities between Tractarian objects and Leibnizian monades. In that case, the distinction between these two concepts of possibility in *Tractatus* does not depend on whether the space metaphor is introduced; instead, it depends on which of the two concepts of space is involved.

1. Introduction

In MS104, a source manuscript for Wittgenstein’s *Prototractatus* (an early draft version of *Tractatus Logico-Philosophicus*), there are two consecutive propositions which, however, seem to be incoherent with each other: [2] “If I can conceive a thing *in* a situation, then I cannot conceive it out of the situation” and [3] “Each thing is, as it were, in a space of possible states of things. I can imagine this space empty, but not the thing without the space” (MS104: 49). It sounds like the former proposition claims that something that is imaginable in a situation, *cannot* be removed from the situation, yet the latter says that it *can*, because only by removing all the possible things in a situation we shall get to the imagination of an “empty” space.

This feeling of incoherence is carried over to *Tractatus*. On the one hand, MS104: 49[3] is authentically retained as “Each thing is, as it were, in a space of possible states of affairs. This space I can imagine empty, but I cannot imagine the thing without the space” (TLP 1972: 2.013). On the other hand, although MS104: 49[2] itself is missing in *Tractatus*, we can easily find a counterpart of it: “If I can imagine objects combined in states of affairs, I cannot imagine them excluded from the possibility of such combinations” (TLP 1972: 2.0121).

Here, the author says even more confidently that objects cannot be removed from the possibility of such contexts. Therefore, one cannot explain away the above incoherence in MS104 by saying that it is only a manuscript for *Prototractatus* and that the confusion is eventually removed in *Tractatus*.

Some philosophers suggest that such a problem shows an ambiguity in the concept of possibility in *Tractatus*. For example, Peach argues that there are two distinct concepts of possibility: the “space metaphor” concept, which compares the possibility of a state of affairs to a spatial point, and the “object-based” concept, which says that the possibility of an object’s occurring in a state of affairs lies in the object itself (Peach 2007: 636). Cerezo also defends a similar view that makes a distinction between the notion of possibility as thinkability and the notion of possibility*, which is revealed by thinking of an empty logical space (Cerezo 2012: 646). In this paper, I will first examine these two concepts of possibility in *Tractatus*, and then compare them with the two different concepts of space.

2. Possibility as an Object-based Concept

At the beginning of *Tractatus*, Wittgenstein makes an ontological claim of the nature of the world: “The world is the totality of facts, not of things” (TLP 1972: 1.1). Here, the word “fact (*Tatsache*)” should be viewed as a technical term, whose definition is to be found immediately in the same page: “fact” means “the existence of states of affairs”, whereas “state of affairs” is defined as “combination of objects (things)” (TLP 1927: 2, 2.01). Thus, in asserting that the world is *not* the totality of things, Wittgenstein is surely not denying the very existence of those individual objects, but only emphasizing that what should be focused on is the combinations of these objects rather than the objects themselves. This ontological claim is further elaborated by the following two propositions: “The world is determined by the facts, and by their being *all* the facts” and “For the totality of facts determines what is the case, and also whatever is not the case” (TLP 1972: 1.11, 1.12). As Black (1964: 33) observes, these sentences reveal that Wittgenstein uses the terms “world” and “fact” in a way that departs from our traditional and ordinary usages.

One of the important theories proposed by early Wittgenstein is the isomorphic structure between the world and the language. It is sketched in the

first several pages of MS104 and fully developed in the published version of *Tractatus*: the world divides into facts and such facts can be grasped by using propositions through logical pictures and their modes of depiction; these propositions represent the existence and non-existence of states of affairs, for which the existence or non-existence of any single item is independent to that of the others. Moreover, just as the elementary propositions are composed by their linguistic constituents, i.e., names, the elementary facts are composed by objects, to which the names refer. In this sense, the one-to-one correspondence between language and the world is twofold: the elementary propositions are linked to states of affairs by their sharing the same logical forms, and the names relate to objects in light of their being part of the larger pictures.

The understanding of “the world” as being the totality of facts, in addition to the linguistic-ontological isomorphic structure, also suggests the importance of incorporating the concept of possibility into the concept of object: “It is essential to things that they should be possible constituents of states of affairs”, and “if a thing *can* occur in a state of affairs, the possibility of the state of affairs must be written into the thing itself” (TLP 1972: 2.011, 2.012). This is because the possibility of the objects’ entering into some configuration is essential in determining the sense of those propositions containing the corresponding names. To make sense of saying some proposition is true, we must be able to imagine when and under what condition it is not true, and thus, in understanding such conditions, given the isomorphic correspondence between language and ontology in *Tractatus*, it requires us to imagine how the objects combine with one another—both in the way they do in this actual world and the ways in which they do not.

Cerezo argues for equating the idea of possibility in this object-based sense to the concept of thinkability: “possible states of affairs are thinkable connections of objects” (Cerezo 2012: 646). For this position, evidence can be found in *Tractatus* as “A thought contains the possibility of the situation of which it is the thought. What is thinkable is possible too” and “If I can imagine objects combined in states of affairs, I cannot imagine them excluded from the possibility of such combinations” (TLP 1972: 3.02, 2.0121). The nature of an object presupposes such possibilities as the “internal properties” of this object, only by grasping which we shall get to know the object: “If I know an object, I also know all its possible occurrences in states of affairs” (TLP 1972: 2.0123).

All of the above suggests that this concept of possibility is established on the basis of objects and their properties. As Cerezo puts it, the only ontological commitment we need for this very conception of possibility “is thus the existence of the objects themselves, which ‘generate’ the whole set of possibilities” (Cerezo 2012: 647).

This ontological commitment gives a realist interpretation of *Tractatus*: the language mirrors the world by having those “fixed” and “unalterable” objects as the referents of names. From the existence of objects, we can not only generate the possible states of affairs, but also the possible worlds: “It is obvious that an imagined world, however different it may be from the real one, must have something—a form—in common with it. Objects are just what constitute this unalterable form” (TLP 1972: 2.022, 2.023). In constructing an imagined world, we have to first think about this real world by breaking the facts down into states of affairs—which are the combinations of objects—and then imagine how these objects can be combined with each other in a different way. As Wittgenstein suggests, if we know these objects, we should be able to do so, and picture to ourselves such imagined worlds.

3. Possibility and Logical Space

The realist interpretation provides an elegant way of understanding the ontological statements in *Tractatus*. However, it is controversial as to whether those Tractarian objects really exist. After all, Wittgenstein himself doesn’t give any concrete examples for those simple objects. What is known to us is only that the ordinary objects (such as cats, mats, brooms...) are not Tractarian objects, because these objects we encounter in everyday life are complex, alterable and something that could not have existed, whereas the Tractarian objects, being the ontological foundation of the world, are “simple”, “unalterable” and “subsistent” (TLP 1972: 2.026, 2.027). Moreover, it is confirmed by Wittgenstein himself in the *Notebooks* that such ordinary names “must disappear on further analysis” and thus the question “Are there simple things?” must be nonsense (Sluga 2012: 111, 114).

Non-realist interpreters such as McGuinness and Ishiguro believe that the Tractarian objects do not really exist. Instead, they argue that the suggested hierarchical structure of the world, which presupposes the existence of simple

objects, is only a “myth”, which is deliberately fabricated to show the nature of our language. Thus, it should be entirely wrong to understand *Tractatus* as providing a metaphysical theory, which is “the very kind of endeavor Wittgenstein explicitly condemned” (Peach 2007: 647). If they are correct, then any endeavor of interpreting the concept of possibility, as well as any other concepts in *Tractatus* that requires the ontological commitment of existence of simple objects, should be problematic.

An alternative interpretation that does not involve this ontological commitment is to understand the concept of possibility in the logical space: “A picture represents a possible situation in logical space” and “A picture contains the possibility of the situation that it represents” (TLP 1972: 2.202, 2.203). Black suggests that here the word “contains” is important, and he emphasizes that “picture ‘contains’ the possibility by actualizing it” (Black 1964: 92). If we compare his comment with Wittgenstein’s famous claim that “A proposition is a picture of reality” (TLP 1972: 4.01), then this reality should be understood as a possible situation that we imagine, which has yet to be actualized. A place where no possible situation has been actualized is empty.

Moreover, one of the central ideas in *Tractatus* says that propositions are truth-functions of elementary propositions, and the truth or falsehood of each elementary proposition depends on whether the corresponding states of affairs exist (TLP 1972: 5, 4.25). As the existence and non-existence of states of affairs are independent from each other, we can draw ourselves a large truth-table and put their corresponding elementary propositions at the top of each of the columns. It turns out that the actual world, which is the totality of the existence and non-existence of states of affairs, can be described by one of the rows of this truth-table, whereas the totality of all possible worlds is to be depicted by the entire truth-table in which all the elementary propositions are contained. Therefore, Cerezo concludes that under this interpretation “*logical space* is what is expressed by the truth-table” and for this space to be empty, it means “there to be no object that occupies any of the positions in the corresponding range” (Cerezo 2012: 649, 654).

4. Two Concepts of Space

It seems that Wittgenstein likes to use the geometric metaphor of logical space: “A proposition determines a place in logical space”, “The propositional sign with logical coordinates—that is the logical space” and “In geometry and logic alike a place is a possibility: something can exist in it” (TLP 1972: 3.4, 3.41, 3.411). In this metaphor, the words “place”, “empty”, and “coordinates” strongly indicate a familiar Newtonian idea of space, which is characterized by Black as “something existing independently of the presence of material bodies and having its own properties that can be stated without reference to such bodies” (Black 1964: 3). But this is not the only way the concept of space can be understood. According to Black, the term “space” was more closely associated with material objects before the seventeenth century. It is the emergence of Cartesian geometry that makes the concept of space become more abstract, because in the Cartesian coordinate system, places in the space are defined as ordered real numbers. In Newtonian mechanics, the positions of physical objects are to be represented by the coordinates, but the coordinates themselves, and thus the whole space, do not need to rely on the existence of those objects. If we understand the Tractarian logical space in this way, then it seems natural to imagine this space as empty.

However, Leibniz rejects the Newtonian idea of space. In a letter to Clarke, he complains, “according to Sir Isaac Newton’s philosophy (for mathematical principles determine nothing in the present case), matter is the most inconsiderable part of the universe. The reason is because he admits empty space besides matter and because, according to his notions, matter fills up a very small part of space...the more matter there is, the more God has occasion to exercise his wisdom and power. This is one reason, among others, why I maintain that there is no void at all” (Leibniz 1989: 322). The objection here is made against the absolute empty space, which is supposed to stand by itself without any material objects in it. In his next letter, Leibniz gives a clearer view on his own notion of space: “As for my own opinion, I have said more than once that I hold space to be something merely relative, as time is, that I hold it to be an order of coexistences, as time is an order of successions. For space denotes, in terms of possibility, an order of things which exist at the same time, considered as existing together, without entering into their particular manners of existing” (Leibniz 1989: 324). Therefore, the concepts of

space and possibility, according to Leibniz, cannot be separated from the existence of individual objects.

In these letters, Leibniz's defending of his position largely relies on his idea of God, which sounds not quite convincing to the twentieth-century logical positivists. But his ontological theory itself may not be alien to Wittgenstein. Although no direct comment to Leibniz has been made in *Tractatus*, there are significant similarities between early Wittgenstein's ontology and Leibniz's work on monadology. Leibniz defines "monads" to be simple substances, which exist as components of the composites, and these simple substances can neither begin nor perish. This idea is clearly echoed by the proposition in *Tractatus* as follows: "Objects, the unalterable, and the subsistent are one and the same" (TLP 1972: 2.027). Moreover, the existence of Leibnizian monads and Tractarian objects are both inferred from logical requirements. For Leibniz, "there must be simple substances, since there are composites; for the composite is nothing more than a collection, or aggregate, of simples" (Leibniz 1989: 213). For Wittgenstein, simple objects must exist because they are the reference of names, which stand at the end of analysis: "If the world had no substance, then whether a proposition had sense would depend on whether another proposition was true" (TLP 1972: 2.0211). Thus, the similarity between monads and Tractarian objects may also be implicitly transferred to the concept of space, and it is not implausible that Wittgenstein accepts, at least to a certain extent, the Leibnizian notion of space, which rejects the idea of an absolutely empty space. In that case, the distinction between the two concepts of possibility in *Tractatus* does not depend on whether the space metaphor is introduced; instead, it depends on which of the two concepts of space is involved.

5. Conclusion

There are two interpretations of the concept of possibility in *Tractatus*: one is the object-based concept, which equates the possibility of states of affairs to the thinkability of configurations of objects, and the other is the so-called "space metaphor" concept of possibility, which relies on the analysis of propositions into elementary propositions that are independent from each other. But there are also two interpretations of the concept of space: the

Newtonian concept, which allows the idea of empty space, and the Leibnizian concept, which is inseparable from the simple objects. It seems that, if we allow the word “space” to be used ambiguously, the two concepts of possibility can be put under the same name as “possibility in the logical space”; however, as the ontological commitments behind these two ideas are quite different, they remain distinct concepts, which create the tension between the propositions of MS104: 49[2] and MS104: 49 [3].

References

Black, Max (1964) *A Companion to Wittgenstein's 'Tractatus'*, New York: Cornell University Press.

Cerezo, María (2012) “Possibility and Logical Space in the Tractatus”, *International Journal of Philosophical Studies* 20 (5), 645-659.

Ishiguro, Hidé (1969) “Use and Reference of Names”, in: *Peter Winch (eds.) Studies in the Philosophy of Wittgenstein*, London: Routledge and Kegan Paul, 20-50.

Leibniz, Gottfried (1989) *Philosophical Essays*, translated and edited by Roger Ariew and Daniel Garber, Indianapolis: Hackett Publishing Company.

Peach, Andrew (2007) “Possibility in the Tractatus: A defense of the old Wittgenstein”, *Journal of the History of Philosophy* 45 (4), 635–58.

Sluga, Hans (2012) “Simple Objects: Complex Questions”, in: José Zalabardo (eds.) *Wittgenstein's Early Philosophy*, Oxford: Oxford University Press, 99-118.

Wittgenstein, Ludwig (1972) *Tractatus Logico-Philosophicus*, translated by D. F. Pears and B. F. McGuinness, London: Routledge and Kegan Paul.

Wittgenstein, Ludwig [Forthcoming] *MS104*, translated by David Stern.

Acknowledgement

I am deeply indebted to Prof. David Stern, who kindly allows me to cite his unpublished English translation of MS104.

Translating Wittgenstein's *Tractatus* Into RDF

Jakub Gomułka (Kraków, Poland)

Abstract

Launched in 2006 by the Wittgenstein Archives at the University of Bergen (WAB), the Wittgenstein Ontology project has led to creation of a knowledge graph collecting the metadata of Wittgenstein's papers, thereby rendering it accessible to both computer algorithms and human scholars. This knowledge base serves as the foundation for several *Nachlass* browsing tools available on the WAB website. Over the years, the WAB team has endeavoured to enrich the existing information structure with representations of the philosophical content found in Wittgenstein's *Nachlass*. However, accurately representing such a content is challenging due to its reliance on context-sensitive concepts, its dynamic nature, the potential for conflicting interpretations, and its tendency to propose its own categorisations of reality.

In my several articles published since 2022, I have tried to outline a possible solution to this problem. However, this solution is very labor-intensive, so when moving to putting my ideas in practice, I decided to narrow its scope exclusively to the *Tractatus Logico-Philosophicus*. In short, the task involves mapping the content of the *Tractatus*' theses onto a format compliant with RDF, which means forcing it into a subject-property-object notation.

The present paper focuses on describing RDF syntactic structures through which I strive to faithfully represent Wittgenstein's language. I have assumed that the representation I create should be as neutral as possible; therefore, the structure I build is not purely conceptual but is rather a form of systematisation of natural language, offering hope for the possibility of applying automated reasoning.

The Wittgenstein Ontology project is a pioneering research effort that combines philosophy and information science. Its main goal is to build a knowledge base about Wittgenstein's works that can be easily used by computer algorithms and searched by human users. The knowledge base is being created based on the technology of Semantic Web (SW), developed within the research paradigm of Knowledge Representation and Reasoning, which belongs to the artificial intelligence research. Since it started in 2006, the project has been led by the Wittgenstein Archives at the University of Bergen, Norway (WAB). So far, the WAB team has finished a part of this knowledge base, which includes metadata of Wittgenstein's published and unpublished writings and their structure. The team has also written several papers on how to represent philosophical ideas within the SW representation technology.

SW consists of standards that facilitate the organisation and sharing knowledge in a way accessible to both humans and machines. According to those standards, knowledge is represented as a collection of uniquely identifiable resources linked by binary relations. In this way, networks of resources linked with relations called knowledge graphs are created. Resources that occur in them include entities (objects within graphs), object properties (binary relations between entities), and data properties (alphanumeric values assigned to entities). The Resource Description Framework (RDF), developed by the World Wide Web Consortium (W3C), outlines the basic principles of SW, using RDF triples (subject, property, object) to structure information. This framework, along with RDF Schema, Web Ontology Language (OWL), and SPARQL (a query language), forms the foundation of SW technology. RDF triples are the building blocks of knowledge graphs. The graphs themselves provide common name spaces for their elements and are themselves identified by unique names, usually URLs under which they are accessible. The function of RDF Schema and OWL is to let developers design hierarchies of classes of elements, as well as hierarchies of properties, and their interrelations. Definitions of classes and properties and subsumptions of elements to classes conform to RDF syntax: they are also triples made of built-in special OWL and RDF Schema meta-properties.

Because of the limitations of the subject-property-object triple structure in capturing diverse propositional forms, several additional features of RDF syntax enhance the expressiveness of graphs. Among these is the concept of the blank node, a type of pseudo-entity that can function as either a subject or an object in a triple without being a distinct resource, meaning it lacks a unique identifier. While it may be assigned a local name within a specific graph for identification purposes, it is also possible for it to remain unnamed. Another interesting feature of RDF syntax is reification, enabling the depiction of a sentence (essentially, an RDF triple) as an object. This object is an instance of a special built-in RDF class named 'Statement.' An object of a type 'Statement' is ascribed three unique object properties: 'subject,' 'predicate,' and 'object.' These properties are assigned the corresponding members of the triple intended for reification. This syntax feature enables to formulate RDF triples about triples (that is, sentences about sentences) allowing for more nuanced and meta-level representations of information.

The SW paradigm is particularly suited for representing well-defined, undisputed knowledge, such as the structure of a collection of documents, linking them to related people, topics, and dates. This approach is exemplified by the Wittgenstein ontology project managed by the WAB, which serves as an automatic index for Wittgenstein's complete works and remarks. However, not all knowledge, especially in philosophy, is this clear-cut. Philosophical texts often encompass vague concepts, dynamism, and conflicting interpretations. The WAB team acknowledges the complexity of accurately representing philosophical knowledge (Mácha et al. 2013). Their initial approaches, which attempted to draw parallels between the structural aspects of computational and philosophical ontologies, such as those implied by the *Tractatus Logico-Philosophicus* (Zöllner-Weber and Pichler 2007), eventually led to the realisation of the need for a novel approach to knowledge representation. WAB's recent research has culminated in the proposal of a conceptual framework centred around three key classes: 'Perspective,' 'Claim,' and 'Concept.' However, this proposal remains primarily theoretical and the WAB 2021 publication concludes with an acknowledgment of the provisional nature of their findings, underscoring the absence of definitive solutions and detailed examples, highlighting the ongoing challenges in the endeavour to accurately represent philosophical knowledge within the RDF structures (Pichler et al. 2021).

In 2022, I commenced a collaboration with WAB to implement a segment of the Wittgenstein ontology related to the philosophical content of Wittgenstein's texts. In several recently published papers (Gomułka 2023a, Gomułka 2023b, Gomułka 2023c, Gomułka 2023d), I have included some specific syntactic solutions that would allow overcoming at least some of the difficulties identified by WAB. First and foremost, I suggested that all constructions directly representing content should be reified and as objects of the 'Claim' class assigned to specific interpretative perspectives. This stems from the fact that the attempt to represent content expressed in natural language within RDF syntax is a translation into a language with significantly lesser expressive capabilities, and thus, potentially a lossy translation, which ultimately constitutes a form of interpretation. This solution opens up the possibility of creating various alternative representations of the same philosophical content, which is especially important in the case of texts interpreted differently by

various researchers. Secondly, my proposal assumed that the content of each sentence in Wittgenstein's *Nachlass* would be represented by one or more instances of the 'Claim' class, while instances of the 'Concept' class will form reified structures assigned to 'Claim' elements. Thirdly, these structures would map the structures of Wittgenstein's sentences (or their functional parts) within the capabilities offered by RDF syntax, including the use of blank nodes. Fourthly, object properties that would occur in these structures should be a kind of meta-properties in relation to the subject-matter content being represented. In other words, 'Concept' instances encoding content elements, which could be named, following Frege, as unsaturated expressions, would always occur in RDF triples in the position of the subject or object. Finally, it would be also necessary to represent various relations occurring between 'Concept' instances themselves, such as being a part of, or being a variant of another instance.

Translating content expressed in natural language into RDF syntactic structures is labor-intensive, making it difficult to envision implementing this for the entire *Nachlass* with the currently assumed workflow: it would simply require too much effort. (This doesn't mean that the process can't be at least partially automated: the current state of large language models (LLM) technology offers some hope in this regard.) For this reason, I decided to limit my project exclusively to the *Tractatus Logico-Philosophicus*. Currently, this project has entered the phase of practical implementation: the first partial knowledge graphs representing the ordinal theses of the *Tractatus* have already been generated and presented to the public.

Before I delve into the more detailed discussion of specific solutions employed during the execution of this project, which I've named Tractarian Ontology, I need to clarify a general issue: *what* exactly is represented in the knowledge graph? Is it the linguistic structure or the conceptual content of the *Tractatus*? The answer is not straightforward because in creating it, I am guided by two seemingly contradictory goals: first, to create a structure made of unequivocal elements that allows for the automatic generation of new information not explicitly entered into the knowledge graph, in line with the SW paradigm; second, to create a representation that is uncontroversial, hence as close to the literal wording of the *Tractatus* as possible. The first goal would suggest a conceptual structure representation, while the second would imply a linguistic

one. Consequently, the approach I've taken is intermediate: I aim to identify formal dependencies between individual content elements of the *Tractatus* while also striving to faithfully reproduce its linguistic structure. An important source of my inspiration is medieval supposition theory (or more precisely, the various medieval theories of supposition, as there was no single dominant one), which attempts a formal approach to linguistic expressions.

Let us start with the structure of the class hierarchy. In my previous texts, I proposed various options for expanding the 'Subject' branch. For example, in 2023d, I suggested it be supplemented with subclasses 'Formula,' 'Ruleset,' and 'Clause.' The first was to group representations of logical and mathematical formulas written in TeX notation, the second's instances were to be rules explicated in the content of the *Tractatus* for processing these formulas, and the third was to group sentence-like expressions that would not be claims affirmed in the text but logical components of such claims. Currently, I consider only the first necessary. 'Ruleset' – if the processing of formulas is to be included at all – should rather be implemented as a part of some external system for generating additional content (e.g. a part of Tractarian Wittgenstein, see below). Meanwhile, 'Clause' instances can simply be instances of 'Claim' that are not assigned to any sentence of the *Tractatus* (contrary to instances of 'Claim' that are assigned and therefore are considered *Tractatus* claims). In place of these classes, I now propose adding three 'Concept' subclasses: 'ConceptVariable,' 'Set,' and 'Quantification.' These classes group instances that play the role of concepts in representations but constitute formal constructions based on the idea of a variable, meaning a representative of an individual or group of individuals falling under a specific concept. In this way, one of the fundamental problems of representing philosophical content, that it usually includes categorisations of reality, is resolved.

The foundation for creating the *Tractatus* representation is its original German text (e.g. Klement 2018), which is why all instances of the 'Concept' class – unless they are formal constructions – are expressions in German. These expressions can be both simple ('Bild') and complex ('Bilder der Tatsachen sich machen'). In the case of simple expressions, the convention of lemmatisation of all parts of speech is adopted, meaning nouns always appear in the nominative singular form without an article, verbs in the infinitive present tense, etc. In the case of complex expressions, internal articles and

grammatical forms of dependent components are preserved, but verbs also take infinitive present tense forms. It should be noted that the 'Concept' class includes all expressions that can function as predicates, thus encompassing simple nouns like 'Welt,' 'Sachverhalt,' etc. as well as verbs such as 'machen,' 'bestimmen,' etc., and adjectives like 'positiv,' 'negativ,' etc.

In addition to German expressions, the 'Concept' class also includes constructions that represent various functions in which a given expression appears within the context of a sentence (or, according to medieval logicians, its suppositions). Let's examine the following example from the *Tractatus*: 'Der Sachverhalt ist eine Verbindung von Gegenständen.' Here, the concept 'Gegenstand' is used to denote a certain number of its designates (not all, which is known from the broader context). In order to express this we need to create an instance of 'ConceptVariable' class; that is, a variable of a concept 'Gegenstand.' Next, we need an instance of 'Quantification' class which is built on a base of that variable and is attributed with a special data property 'plural' which marks its numerosity. Since 'Quantification' is a subclass of 'Concept,' that instance is also a kind of 'Concept' that can occur as a complement in a complex concept 'Verbindung von Gegenständen.' Take the next example sentence from the *Tractatus*: 'Es ist dem Ding wesentlich, der Bestandteil eines Sachverhaltes sein zu können.' Here, we have the concept 'Ding,' which appears in the singular form, but from the context, we understand that the sentence refers to each *designatum* of this concept individually. In this situation, we use an instance of 'Quantification' built on the concept variable 'Ding', to which the data property 'forEach' is assigned, indicating this status. Take yet another example, the well-known first thesis of the *Tractatus*: 'Die Welt ist alles, was der Fall ist.' Here, we have a concept 'Fall sein.' This concept is used in a quite specific way: the entire phrase 'alles, was der Fall ist' denotes one complex entity composed of individual instances, each of which being the case. This represents a rather complex syntactic structure based on the concept variable 'Etwas.' The word 'Etwas' appears multiple times in the *Tractatus*, but not in thesis 1, so this is something I introduced artificially. Since the word has no technical employment in the *Tractatus*, I decided to use it to denote anything, any individual. So, on the concept variable 'Etwas,' we build an instance of 'Quantification' with the data property 'general,' indicating that it refers to all designates simultaneously. The final layer of this

construction is an instance of the class 'Set' (also a subclass of 'Concept'), to which, through the object property 'quantification,' this instance of 'Quantification' is assigned, and also, through the object property 'satisfies,' the concept 'Fall sein.' In other words, the set is defined as containing all designates of the concept 'Etwas,' essentially all individuals, which satisfies the predicate 'Fall sein,' making it a construction equivalent to the expression 'alles, was der Fall ist.'

Connectors of 'Concept' instances (which include German expressions as well as instances of 'Set,' 'Quantification,' and 'ConceptVariable' created based on these expressions) which link them into RDF triples are special meta-properties, which I name in English. These essentially belong to the same category as the already mentioned meta-properties 'quantification,' 'satisfies,' and others used to create complex constructions belonging to the 'Concept' subclasses. While those could be considered as predicate-forming functors, these are sentence-forming functors. I aim to keep the lists of both types of functors not too lengthy, yet before the project's completion, they cannot be closed. An example of a sentence-forming meta-property I use in the representation of theses 1, 2, 3, 4, 5, and 6 is the object property 'equivalent.' Essentially, this clarifies the conjunction 'ist' found in these sentences, which, obviously, can serve many different functions. In other words, this is a clear example that translating the *Tractatus* into RDF is an interpretation: I interpret these backbone Tractarian sentences as a series of definitions.

As was mentioned earlier, RDF triples that directly represent the content of the *Tractatus* undergo reification: their components become values of the properties 'subject,' 'predicate,' and 'object' assigned to instances of the built-in RDF class 'Statement.' These instances are, in turn, assigned as values of the property 'structure' to instances of the class 'Claim.' These, in turn, may or may not be assigned as values of the property 'hasPart' to instances of the class 'Sentence,' which belongs to another (already completed) branch of the class hierarchy in the Wittgenstein ontology. At the same time, 'Claim' instances are assigned to specific instances of the class 'Perspective,' indicating that individual reified representations belong to a single view of the *Tractatus* (instances of 'Perspective' can, in turn, be assigned to instances of the class 'Person' present in Wittgenstein ontology).

These are the foundational principles behind the Tractarian Ontology project. It is immediately apparent that the final structure of representation of the Tractarian philosophical content will be highly complex and generally not user-friendly. While it will be possible to create SPARQL queries that yield meaningful results, these queries will need to be lengthy and hardly legible. For this reason, the Tractarian Ontology is a project developed from the outset with a broader system in mind, tentatively named Tractarian Wittgenstein, intended to serve as an advanced interface for this ontology. There isn't enough space here to present the detailed premises of this system, but it's worth noting that it is intended to include components functioning as algorithmic processors in a classical style with explicitly stated processing rules (these will be responsible for generating SPARQL queries and interpreting their results) as well as a module based on LLM technology that will enable communication with the user in natural language.

The role of Tractarian Wittgenstein will be to provide answers to any reasonable question posed in natural language concerning the content of the *Tractatus Logico-Philosophicus*. Once this system is developed, it will enrich the toolkit available to the public on the WAB website.

References

Gomułka, J. (2023a) "Artificial Intelligence applied to philosophy: a contribution to the Wittgenstein Ontology project", *Ethos : Kwartalnik Instytutu Jana Pawła II KUL*, 36 (3), 207–234.

Gomułka, J. (2023b) "Technologia sieci semantycznej w zastosowaniu do humanistyki: przypadek Traktatu logiczno-filozoficznego Wittgensteina (Semantic web technology as Applied to the humanities: the case of Wittgenstein's Tractatus Logico-Philosophicus)", in: A. Jonkisz, J. Poznański, J. Koszteyn (eds.), *Zrozumieć postrzeganie i pojmowanie człowieka oraz świata : prace dedykowane Profesorowi Józefowi Bremerowi SJ z okazji 70-lecia urodzin (To understand perception and comprehension of the human and the world : papers dedicated to Professor Józef Bremer S.J. on the occasion of his 70th birthday)*, 419–439. Kraków: WN AIK.

Gomułka, J. (2023c) "Towards a computational ontology for the philosophy of Wittgenstein: representing aspects of the Tractarian philosophy of mathematics", *Analiza i Egzystencja*, 63, 27–54.

Gomułka, J. (2023d) "The Wittgenstein Ontology: Representing the Tractatus", in: E. Heinrich-Ramharter, A. Pichler, F. Stadler, J. Wang-Kathrein (eds.), *100 Years Tractatus: Contributions of the 44th International Wittgenstein Symposium (6–12 August 2023, Kirchberg am Wechsel)*, 235–242. Kirchberg am Wechsel: ALWS.

Klement, K.C. (2018) *Ludwig Wittgenstein: Tractatus Logico-Philosophicus Side-by-Side-by-Side Edition*. Amherst: University of Massachusetts (<https://people.umass.edu/klement/tlp/> – access: 21.02.2024).

Mácha, J., Falch, R., Pichler, A. (2013) "Overlapping and Competing Ontologies in Digital Humanities", in: F. Tomasi, F. Vitali (eds.), *DH-CASE '13: Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (10 September 2013, Florence, Italy)*, ACM.

Pichler, A., Fielding, J.M., Gangopadhyay, N., Opdahl, A.L. (2021) "Crisscross ontology: Mapping concept dynamics, competing argument and multiperspectival knowledge in philosophy", *Quaderni di "Filosofia"*, 2, 59–73.

Zöllner-Weber, A., Pichler, A. (2007) "Utilizing OWL for Wittgenstein's *Tractatus*", in: H. Hrachovec, A. Pichler, J. Wang (eds.), *Papers of the 30th International Ludwig Wittgenstein Symposium (5–11 August 2007, Kirchberg am Wechsel)*, 249–251. Kirchberg am Wechsel: ALWS.

Acknowledgement

I am very grateful to the University of Bergen and the Wittgenstein Archives for inviting me to a research visit and for creating excellent conditions for my work. In particular, I would like to thank the director of WAB, Alois Pichler, for his invaluable help and support. I also owe words of thanks to Simo Säätelä, Jan Ole Bangen, Filippo Mosca, Andreas Lothe Opdahl, Krzysztof Kutt, Claus Huitfeldt, Heinz Wilhelm Krüger, Rune Jensen Falch, as well as to all other participants of the seminars organised by WAB in February 2024 for their illuminating remarks and questions.

Fundamentality, Existence, Totality: On Three Notions of Reality and the Landscape of Metaphysics

Dustin Gooßens (Bochum, Germany)

Abstract

Metaphysics is, historically as well as systematically, mostly taken to be the inquiry into reality, insofar it is considered to be: (1) the *totality* of everything there is; (2) of everything that *exists*; or (3) what is *fundamental*. This paper sets out to analyze the relation between all three metaphysical core notions and sketch the landscape of metaphysical theories that emerges from it. Taking *The Fundamental*, *The Existent*, and *Totality* to be the domains corresponding to each metaphysical object of inquiry, it is argued that they stand in the set-theoretical relations of: $The\ Fundamental \subseteq The\ Existent \subseteq Totality$. This general structure allows for a plurality of more detailed structures when we differentiate between the ones that treat at least two notions as extensionally equal and those that take them to be proper subsets. Furthermore, reductive and inflationary strategies between equating two of the notions will be differentiated, allowing for a detailed sketch of the landscape of metaphysical theories and their views on the nature of reality. In closing, it is argued that the metaphysical structure allowing for most differentiation in metaphysical status is to be preferred as it allows to adequately capture metaphysical disputes, constituting evidence for the importance of each metaphysical notion and a short plea against their conflation.

1. Introduction

Historically as well as systematically, the following three notions have proven to be good candidates for what might be taken to be the central topic of metaphysics:

1. *Totality*: Metaphysics as the most general science
2. *Existence*: Metaphysics as ontology, i.e., the science of everything that exists
3. *Fundamentality*: Metaphysics as the science of the fundamental layer of reality

While debates in metaphysics mostly focus on either discussing which of these conceptions is the right one, or on questions that emerge within one of the three, this paper sets out to take a step back and look at the relations between all three notions. It will proceed as follows: Firstly, possible relations between the three notions will be discussed and one argument for each relation will be developed. Secondly, the landscape of possible metaphysical theories that emerges from this general structure will be sketched. The decisions leading to these different structures will be connected to two long-lasting debates in

metaphysics. This structure will then be refined further by distinguishing between *reductive* and *inflationary* strategies. Thirdly, a plea for a specific structure (the "most permissible metaphysical structure") will be brought forward.

2. The General Metaphysical Structure

Let us take *Totality*, *The Existent*, and *The Fundamental* (with capital letters) to be domains corresponding to the objects of study of the different notions of metaphysics. I do not want to commit to any in detail description of what each domain encompasses or what exactly the domains themselves are (it is probably most natural to assume they are *sets*) as the aim is to make a general point neutral to (almost any) specific metaphysical theory.

The general relation, in which all three domains stand to one another, is the following:

$$\textit{The Fundamental} \subseteq \textit{The Existent} \subseteq \textit{Totality}$$

In natural language: what is fundamental is a part of what exists, which is, in turn, a part of the totality of everything there is.

Let me argue for each relation separately. That *The Existent* is a subset of *Totality* straightforwardly follows from the definition of *Totality*. If this would not be the case, there would exist something that is not part of *Totality*, undermining that the *Totality* in question is truly the absolute totality of *everything*.

That *The Fundamental* is a subset of *The Existent* follows from the following argument: If there were something that is fundamental but not existent, then either it is the ground of something existent, which contradicts the metaphysical principle of *ex nihilo nihil fit*. Or, it does not ground anything, but then it is neither explanatory nor existent, which makes the entity in question theoretically obsolete.

There are at least two counter-examples that might come to mind. Firstly, some authors take there to be fundamental negative facts (see, e.g., Amijee 2021). The response to this is rather laconic, as these negative facts are also

taken to *obtain*, which is, in turn, an existence predicate for facts (Moltmann 2020: 327), so these are no actual counter-examples. The other counter-example that might come to mind can be seen in theories that take *Nothingness* to be the ultimate ground of reality. This has, e.g., been proposed by Priest (2021) (with references to Heidegger and Nishida), who claims (bold face) **nothing** to be the ultimate ground of everything for the following reason:

Something (*g*) being an object depends on its being distinct from **nothing**. If *g* were the same (in ontological status) as **nothing**, it would not be an object, since **nothing** is not an object. (Priest 2021, 19)

While Priest himself claims **nothing** to both be an object and not be an object (hence taking the ground of reality to be contradictory in nature), one might not want to admit **nothing** to *exist* but still to be fundamental. However, even such a version would not constitute a counter-example to the proposed structure, as Priest conceptualizes **nothing** as the fusion of the empty set, which results in the fact that "[t]he empty fusion [...] is a part of everything" (Priest 2014, 98). As this **nothing** thus is a subset of *The Existent* as well, it does not constitute a counter-example either.

As I take the general structure to be widely acknowledged and intuitive, I hope this suffices as a justification and defense of it. I will now turn to sketching the landscape of metaphysical theories that emerges from the general structure.

3. A Landscape of Metaphysics

Since the relation between each of the domains corresponding to the three notions of metaphysics still allows for the two options of either being in fact *extensionally equivalent* or a *proper subset*, the following four possible combinations emerge:

1. *The Fundamental* = *The Existent* = *Totality*
2. *The Fundamental* \subset *The Existent* = *Totality*
3. *The Fundamental* = *The Existent* \subset *Totality*
4. *The Fundamental* \subset *The Existent* \subset *Totality*

It is quite standard nowadays to compare different ontological theories by distinguishing between a general and not ontologically loaded domain of quantification and a subset of this domain, which is interpreted to be the extension of an existence predicate. Let us call the general domain " T " and the existence predicate " E ". Normal ontological disputes regarding questions of the form "Do dinosaurs exist?" can be seen as just concerning whether the entities in question are to be found just in T or also within the extension of E (Azzouni 2021: 179). We can now enrich this model by including a fundamentality predicate " F ", which, according to the proposed general structure of metaphysics, is to be seen as a subset of E .

Whether F and E are coextensional is essentially the question of whether we take ontology to be flat or structured in the sense of Schaffer (2009: 355). If we claim *The Fundamental* to be equal to *The Existent*, then there is no further structure within ontology, and it becomes *flat*. If we take *The Fundamental* to be a proper subset (be it even the empty set in the case of Priest, or, for instance, in cases of a 'gunky ontology' (Tahko 2018: 239-240) where there is assumed to be an infinite chain of more fundamental layers that never terminates in a truly fundamental one), then our metaphysical picture is *structured* via relations of metaphysical dependence and along degrees of fundamentality.

Whether E and T are coextensional in the sketched framework now primarily becomes a question of whether we take quantification to be ontologically loaded or not. If we take them to be coextensional, then everything in the range of the quantifiers will also exist. If we take it to only be a proper subset, then we will partially quantify over things within the anti-extension of the existence predicate. Within the sketched formal model, these decisions then correspond to either endorsing the restrictions " $F = E$ " or " $E = T$ " in addition to the general restrictions or explicitly not doing so.

What these structures so far tell us is how many different domains there are and their extensional relations. So, it is clear that the structure equating all notions has one domain, the two structures that equate only two of the notions have two, and what I will call "the most permissible structure" has three extensionally distinct domains.

However, this does not yet tell us what kinds of entities are taken to be part of each domain in sufficient detail. To get clearer on this, let me introduce a differentiation between two strategies of equating two notions. The first option is what one might call a *reductive* strategy. This consists in (conceptually) shrinking the more encompassing domain down to the lesser one. This happens, e.g., when authors claim that what *truly* exists *just is* *The Fundamental* (in favor of "F = E") or all there *really* is, is *The Existent* (in favor of "E = T"). However, one could also follow an *inflationary* strategy and extend the smaller domain to the size of the more encompassing one. (The difference between both strategies points towards an explanatory asymmetry between both sides. These asymmetric 'just is'-statements are also briefly mentioned, although not endorsed, by Rayo (2013: 5)). In the following, the order in which the domains appear next to the equality symbol will indicate this asymmetry. The domain to the left is taken to be inflated or reduced to the extension of the right one, imitating the natural language intuition that there is a difference between saying "*The Existent* just is the *The Fundamental*" and "*The Fundamental* just is *The Existent*".

The differences between reductive and inflationary strategies do not become apparent when just looking at the abstract structure, but they appear when considering translations between statements made within different structures. This is easiest to see when looking at all three possibilities in the most permissible structure and how these would translate into more restrictive ones. So, let us assume that in the most permissible structure we can speak of three objects "*a*", "*b*", and "*c*". Object *a* is fundamental, object *b* is non-fundamental and existent, and *c* is in the domain of quantification but non-existent. Depending on whether a reductive or inflationary approach to the equality in the structure '*The Fundamental* = *The Existent* \subset *Totality*' is chosen, object *b*'s status will change. On the reductive approach, it will be in domain *T*, but not in E = F, while on the inflationary approach, it will be in E = F as well. *Mutatis mutandis*, the same goes for the equation of *The Existent* and *Totality*. Though one important detail in this case is that objects can, on a reductive treatment, fall out of even the largest domain *T*. If we take the reductive stance that we can only quantify over what exists, object *c* will be outside of *T*. One can adopt some version of a semantics for free logic to account for ways of still meaningfully speaking about them (see, e.g., Nolt 2021; Rami 2020), though

oftentimes authors endorsing this reductive strategy deliberately do not want to admit meaningful speech about such nonexistents. So, if a reductive strategy is adopted, objects, which are in the most permissible structure considered to be only in the larger of the two domains, get 'pushed out' of the smaller domain and into the next greater one. If there is no such domain, then it gets 'pushed out' of the model completely. While an inflationary strategy achieves the opposite to 'pull in' the object in question into the inner domain. In this sense, from the standpoint of the most permissible structure, we can determine the strategy chosen by a different metaphysical system by asking for the right translation of objects like b and c into the structure in question. If the object is taken to be part of the equated domains, it is an inflationary strategy. If it is 'pushed out', the pursued strategy is a reductive one.

Accounting for this asymmetry, the landscape of metaphysics now becomes a bit more complicated (parentheses are added for disambiguation and easier readability):

1. *The Fundamental* \subset *The Existent* \subset *Totality* (Most Permissible)
2. *Totality* = *The Existent* = *The Fundamental* (Fully Reductive)
3. *The Fundamental* = *The Existent* = *Totality* (Fully Inflationary)
4. *Totality* = (*The Fundamental* = *The Existent*) ((F-E)-Inflationary/(E-T)-Reductive)
5. (*The Existent* = *The Fundamental*) \subset *Totality* ((F-E)-Reductive)
6. (*The Fundamental* = *The Existent*) \subset *Totality* ((F-E)-Inflationary)
7. *The Fundamental* \subset (*Totality* = *The Existent*) ((E-T)-Reductive)
8. *The Fundamental* \subset (*The Existent* = *Totality*) ((E-T)-Inflationary)

Note that a (F-E)-Reductive/(E-T)-Inflationary structure is not possible. The two equality signs tell us that there is only one domain. However, being (F-E)-Reductive would mean for a non-fundamental existing object to not be part of the domain, while being (E-T)-Inflationary would mean for a non-existent object to be part of the domain. This, however, contradicts the general structure (*The Fundamental* \subseteq *The Existent* \subseteq *Totality*) and is thus impossible.

4. A Plea for The Most Permissible Structure

It is important to see that the type of questions we face now is of a different kind than the rather ordinary ontological questions about the existence of dinosaurs or electrons. The two choices of how to conceptualize the relations between *The Fundamental*, *The Existent*, and *Totality* are conceptual or theoretical choices about the imposed framework, rather than choices within a given framework (one might think of them rather as practical/external questions than internal questions in the sense of Carnap (1950)). What we mean when we equate either *The Fundamental* and *The Existent* or *The Existent* and *Totality* is usually not, that contingently it turns out that everything that exists is fundamental, or that it just happens to be the case, that everything there is also exists. Rather, it is a choice regarding which theory one wants to adopt *before* engaging in any object-level inquiry. These types of metaphysical disagreement can thus rather be seen as metalinguistic negotiations (Thomasson 2017).

But exactly this difference between object-level disagreement about the metaphysical status of certain entities and metalinguistic disagreements about what it means to be fundamental, existent, or how quantification should be employed, is in the easiest way reliably drawn when referring to the most permissible structure. Let me illustrate this by way of a simple example. Take there to be a disagreement between a metaphysician endorsing an (F-E)-Reductive structure and one endorsing an (F-E)-Inflationary one. What they disagree about, is then the ontological status of a certain class of entities. From the perspective of the most permissible structure, it becomes evident that the class of entities the disagreement is about is the class of non-fundamental existents. However, if both participants of the disagreement firmly stick to only using their own structure, the disagreement will seem like an object-level disagreement about the existence of chairs, persons and the like, while it should be seen as a disagreement about metaphysical theories. An example more often discussed in the literature is the comparison of ordinary existence assertions and more serious ones devised in the metaphysician's artificial language "Ontologese" or "in the ontology room" (Dorr 2005; Sider 2011). Ontologese corresponds to the fully-reductive structure, as it only speaks and quantifies over fundamental entities, while I take more ordinary metaphysics to resemble a (F-E)-Inflationary/(E-T)-Reductive structure. Several difficulties

for adopting and translating Ontologese from an ordinary speaker's perspective have been pointed out in the literature (Hewitt 2018; Wrigley 2018). What seems like a natural response from the proposed landscape of metaphysical structures is that the dispute between both languages becomes understandable as a metalinguistic negotiation from the point of view of the most permissible structure. In fact, proponents of Ontologese usually introduce their language by using the distinctions that only make sense from a standpoint that distinguishes between all three metaphysical core notions (see, e.g., the "Instructions for Introducing Ontologese" in Sider (2009: 415)). This points towards the fact that we need to differentiate between all three metaphysical core notions to explicate the different metaphysical structures and frame disagreements between them in understandable and correct ways.

Summarizing my plea for the most permissible structure: It is the most fine-grained structure just in virtue of not conflating any of the three metaphysical notions, allowing it to speak about every kind of possible metaphysical status of a given entity. This also explains how it can be used to show the true nature of certain metaphysical disputes, and why it is regularly used by authors, while then turning around and arguing against it. I take this to be an indication of the importance of all three notions in metaphysical theorizing and an argument for keeping them separate.

References

- Amijee, Fatema (2021) "Something from Nothing. Why Some Negative Existentials are Fundamental", in: Sara Bernstein and Tyron Goldschmidt (eds.) *Non-Being. New Essays on the Metaphysics of Non-Existence*, Oxford: OUP, 50-68.
- Azzouni, Jody (2021) "Defending the importance of ordinary existence questions and debates", in: Ricki Bliss and JTM Miller (eds.) *The Routledge Handbook of Metametaphysics*, New York: Routledge, 171-183.
- Carnap, Rudolf (1950) "Empiricism, Semantics and Ontology", *Revue Internationale de Philosophie* 4(11), 20-40.

Dorr, Cian (2005) "What we disagree about when we disagree about ontology", in: Mark Eli Kalderon (ed.) *Fictionalism in Metaphysics*, Oxford: OUP, 234-86.

Hewitt, Simon (2018) "If an ontologist could speak we couldn't understand him", *Inquiry: An Interdisciplinary Journal of Philosophy* 61(4), 444-460.

Moltmann, Friederike (2020) "Existence Predicates", *Synthese* 197(1), 311-335.

Nolt, John (2021) "Free Logic", in: Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), URL = <https://plato.stanford.edu/archives/fall2021/entries/logic-free/>.

Priest, Graham (2014) *One. Being an Investigation Into the Unity of Reality and of its Parts, Including the Singular Object Which is Nothingness*, Oxford: OUP.

Priest, Graham (2021) "Nothingness and the Ground of Reality. Heidegger and Nishida", in: Sara Bernstein and Tyron Goldschmidt (eds.) *Non-Being. New Essays on the Metaphysics of Non-Existence*, Oxford: OUP, 17-33.

Rami, Dolf (2020) "Single-domain free logic and the problem of compositionality", *Synthese* 198(10), 9479-9523.

Rayo, Agustín (2013) *The Construction of Logical Space*, Oxford: OUP.

Schaffer, Jonathan (2009) "On what grounds what", in: David Manley, David J. Chalmers and Ryan Wasserman (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: OUP, 347-383.

Sider, Theodore (2009) "Ontological Realism", in: David Manley, David J. Chalmers and Ryan Wasserman (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: OUP, 384-423.

Sider, Theodore (2011) *Writing the Book of the World*, Oxford: OUP.

Tahko, Tuomas E. (2028) "Fundamentality and Ontological Minimality", in: Ricki Bliss and Graham Priest (eds.) *Reality and its Structure*, Oxford: OUP, 237-253.

Thomasson, Amie L. (2017) "Metaphysical Disputes and Metalinguistic Negotiation", *Analytic Philosophy* 58(1), 1-28.

Wrigley, Wesley (2018) "Sider's Ontologese Introduction Instructions", *Theoria* 84(4), 295-308.

Acknowledgement

Many thanks to Niklas Kurzböck, Jan Köpping and Svitlana Stupak for helpful comments on a previous version of this paper as well as to Andrew Tedder for many constructive discussions.

Quantifiers, Grounds, and Ontological Commitment

Martin Grajner (Dortmund, Germany)

Abstract

Extended Abstract. The quantifier approach to ontological commitment has recently acquired a bad reputation. In this paper, I argue that the quantifier approach can withstand the main objections that critics have levelled against it.

I first wish to sketch a preliminary version of my account. The preliminary version is able to cover the so-called explicit commitments of a given statement. By “explicit commitments,” I mean the ontological commitments that form part of a statement’s subject matter. After that, I will introduce a refined version of my account that captures the *implicit* commitments of some statement. Very roughly, the implicit commitments of “p” are those commitments that “p” does not wear on its sleeves.

Here is the preliminary version of my account:

- (OC) “p” is ontologically committed to Fs iff (i) “p” logically entails “There are Fs”; (ii) there is at least one F in the domain D over which the quantifier contained in “There are Fs” ranges; and (iii) “p” is about an F or Fs.

Condition (i) and (ii) of this account claim that “p” carries ontological commitment to Fs in case “p” logically entails a true quantified statement in the form of “There are Fs,” and there is indeed an F in the domain over which the quantifier in “There are Fs” ranges. This component of the account preserves the main tenet of the quantifier account, according to which ontological commitment is at least in part a function of true existential quantification.

Note that both conditions of the account (OC) are compatible with the fact that true existential quantification and ontological commitment might come apart. The present account claims only that a true statement in the form of “There are Fs” carries ontological commitment to Fs in case there is indeed an F in the domain over which the quantifier ranges. Consider true existentially quantified statements that do not carry any ontological commitments, such as “There are tables” (in case mereological nihilism is true) or “There is chance that it might rain today.” Across a series of papers, Thomas Hofweber (2009, *Ambitious, yet Modest Metaphysics*. In Chalmers, D. et al., eds., *Metametaphysics*, Oxford: OUP, 260–89; 2016. *Ontology and the Ambitions of Metaphysics*. New York: OUP) has defended the view that natural language quantifiers are polysemous. On his proposal, natural language quantifiers have what he calls an “internal” reading as well as an “external” reading. On the external reading, “There are Fs” is a claim about the objects in a given domain over which the quantifier ranges. The statement “There are Fs” will be true on this reading only if the domain indeed contains some thing that is an F. The internal reading, conversely, is neutral with respect to the question whether there is an entity in the domain over which the quantifier ranges. Very roughly, the internal reading corresponds to a substitutional interpretation of the quantifier.

Armed with the distinction between these two readings, a proponent of the quantifier account of ontological commitment can deal with the allegedly problematic cases. If we accept the view that quantifiers have an internal reading, we incur no commitment to tables (if

mereological nihilism is true) in case we utter “There are tables.” Moreover, the truth of this statement also does not require that clauses (i) and (ii) of the present account (OC) are satisfied, which would entail commitment to composite objects.

However, (OC) as it stands does not preclude that every true negative existential carries commitment to anything whatsoever. Take as an example “There are no blue swans.” Let us suppose that the whole domain over which the quantifier ranges contains no blue swans. If that is the case, then “There are no blue swans” will carry commitment to *any* F of that domain—for example, penguins, cars, or numbers. However, it seems counterintuitive that every true negative existential will carry these commitments.

I think that cases such as this one can be excluded by the addition of condition (iii). The idea is that “p” needs to be *about* an F or Fs if it is to carry that particular commitment. On this proposal, a statement “p” only carries commitment to blue swans in case it is wholly or partly about blue swans (and the remaining conditions are satisfied). How should we cash out the notion of aboutness or subject matter? Aboutness is a hyperintensional phenomenon. For present purposes, I assume that the notion of aboutness or subject matter is intelligible and coherent and that we have some firm grip on it when we apply it to particular cases.

As I have already mentioned above, (OC) does not do justice to all the desiderata for an account of ontological commitment. To take one example, the statement “{Socrates} exists” carries commitment only to {Socrates} according to (OC), but not to Socrates. The reason why (OC) does not capture any implicit ontological commitments to Socrates is that “{Socrates} exists” is not *about* Socrates. This worry can be dealt with by adding a further condition (iv) to (OC).

(OCG) “p” carries commitment to Fs iff “p” either carries commitment to Fs according to (OC) or (iv) at least one F is involved in partially or fully grounding p.

By invoking condition (iv), the present account becomes able to handle implicit ontological commitments. For example, “{Socrates} exists” on this view carries a commitment not only to the existence of {Socrates}, and therefore to sets, but also to Socrates (and, for that matter, to individual objects or persons) as Socrates grounds the existence of {Socrates}.

Grounding without Essence

Zhiwei Gu (Shanghai, China)

Abstract

This paper argues against the essentialist model of grounding proposed by Fine, which holds that grounding relationships are explained by appealing to the essences of the grounded facts. I contend that grounding, as a metaphysical form of explanation, does not necessarily depend on an essentialist model. Some grounded facts lack a clear-cut essence or any essence at all, in which cases the essentialist model cannot adequately explicate the grounding relationship. As an alternative, I propose a pluralistic model where recognition-related contingent properties, in addition to essences, can ground explanations of grounding relationships. To answer why A is grounded in B, recognition-related properties of B are required - properties which are recognized as relating to A. Crucially, these properties need not be essences of A. They can be contingent features. This pluralistic model accommodates grounding for vague facts like "party-hood" where the recognitional properties like "reveling" ground why an event counts as a party. But it also handles simple cases involving logical and mathematical facts. By permitting contingent recognitional properties as grounds, my model avoids the limitations of the essentialist approach while still capturing the metaphysical explanatory role of grounding.

1. Introduction

A number of metaphysicians (Fine, 2012; Rosen, 2010; Schaffer, 2019; etc.) have recently advocated for a relation of metaphysical grounding, called "in virtue of", as a form of metaphysical explanation, in addition to causal explanation. Unlike diachronic causal explanation, which operates in a forward temporal direction, grounding operates through a constitutive, synchronic form of determination from the more fundamental to the less fundamental. Several examples may elucidate this notion:

- (1) The fact that the ball is red or green obtains in virtue of the fact that it is red.
- (2) The fact that the sky is blue obtains in virtue of the fact that it is cerulean.
- (3) The fact that there is a party obtained in virtue of the facts concerning Tom, Jerry, Alice, etc. reveling.

Taking (1) as an illustrative case, the disjunctive fact is grounded in its true disjunct; that is, the ball's being red metaphysically explains why the disjunctive fact obtains. What accounts for the explanatory relationship between the grounded fact and its grounds? Kit Fine (2012, p. 76) proposes an

essentialist account, suggesting “the particular explanatory connection between the fact C and its grounds may itself be explained in terms of the nature of C.” It means that the explanatory power comes from the nature of the ground that constitutively relates to C. This essentialist view finds support from Rosen (2010), Dasgupta (2014), and others.

While I concur that there exists this “in virtue of” form of metaphysical explanation, I depart from the essentialism-dependent model, which grounds the explanatory connection in the nature or essence of the explained fact. Instead, I propose a plural explanatory conception of grounding rooted in the use of concepts.

2. Fine’s essentialist model

Let us now examine why Kit Fine combines statements of nature or essence with statements of ground. He asserts, “Given that the fact F is grounded in the facts G1, G2, ..., then it lies in the nature of the fact F (or of the items that it involves) that it should be so grounded given that the facts G1, G2, ..., do indeed obtain.” (Fine, 2012, p. 74) For instance, on this view, the grounding relation between the fact that the sky is blue and the fact that it is cerulean lies in the nature of the former. However, there is an issue with this position. Although it is necessarily true that something’s being cerulean entails its being blue, it does not lie in the nature of the fact that the sky is blue that this fact is grounded in the sky’s being cerulean. For being cerulean is merely one specific instance of being blue rather than the essence of blueness itself.

Fine attempts to circumvent this difficulty by generalizing the statement of the grounding because the generalized statements of the grounding purportedly hold in virtue of the essence of the grounded fact. For example, we might say that the fact that the sky is blue is, for any blueish color x, grounded in the fact that the sky is x (Fine, 2012, p. 75).

Fine’s approach to combining statements of essence with statements of ground is ingenious. He is also judicious in approaching the concept of essence, as without a robust and lucid notion of essence, it would be impossible to have a sound conception of grounding if the grounding connection is truly grounded in the essence of the grounded fact. Fine (1994) proposed a definitional

account of essence but rejected a modal account. Let us first examine the modal account: “an object is taken to have a property essentially just in case it is necessary that the object has the property if it exists”(1994, p. 4). This definition suggests that the essence of an object can be sufficiently characterized by a modal fact. However, this account has some counterintuitive implications. Suppose the existence of Socrates entails the existence of the singleton Socrates. It is necessarily true that Socrates belongs to the singleton Socrates if Socrates exists. According to the modal account, therefore, Socrates essentially belongs to the singleton Socrates. But do we genuinely think that it is part of Socrates's essence that he belongs to the singleton Socrates? This seems irrelevant to Socrates's essence. This is also why Fine insisted that the grounding connection is grounded in the essence of the grounded rather than the essence of the grounds. In Socrates's case, what is grounded is the singleton Socrates rather than Socrates himself.

Another objection Fine mentioned is this: it is necessary that the conjunction of all necessary truths holds if Socrates exists. Of course, it cannot be true that among Socrates's essences is that $2+2=4$. One might propose adding a condition of relevance. But counterexamples are still readily constructed. What we need only do is construct necessary truths which involve Socrates but do not state Socrates's essences, such as the singleton Socrates. These objections are fatal to the modal account. Hence, following Fine, grounding theorists (Dasgupta, 2014; Goff, 2019; Rosen, 2010) favor the definitional account.

What is the definitional account? As Fine puts it, “just as we may define a word, or say what it means, so we may define an object, or say what it is.” (1994, p. 2) They are “not merely parallel but are, at bottom, the same.”(1994, p. 13) So the definitional account of essence involves a definition of an object. To understand how essence and definition connect, it is necessary to know what is involved in giving a definition. Fine examines an empirical view that a word means what it does. Accordingly, “bachelor” is correctly defined as “unmarried man” due to its use. However, this view implies that a word is a mere sequence of letters. He further examined a “thicker” view, which sees a word as a combination of a sequence of letters with its meaning. Under this conception, a definition would state an essential property of the word. But he believes that “there is a deeper connection between definition and the formulation of essence” (Fine, 1994, p. 13), which is dropped out in the

thicker view. He claims that an appropriate way of specifying the meaning of a word is by providing some account of the meaning's essence. Thus, saying what it means is just saying what it is.

3. The problem of the essentialist model

If this definitional account of essence works, then it will be a satisfactory candidate to explain the grounding relation. It does work in areas such as logic, mathematics and the like. For example, $p \# q$ is grounded in p , which can be explained by the essence of the disjunction $p \# q$; that is, $p \# q$ is true just in case p or q is true. The reason why the essentialist model works naturally in these areas is because in logic and mathematics, most of their concepts can be clearly defined.

However, a world of concepts in our everyday life cannot be defined in as rigid a way as logical or mathematical concepts. Let me take Goff's favorite example, party-hood, to illustrate this point (2019).

(4) the fact that there is a party is grounded in the fact that Tom, Jerry, Alice, etc., are reveling.

Following Fine's account as Goff does, the grounding relation is essentially explained by the essence of party-hood. People's reveling is supposed to be such an essence that once it is satisfied, the fact that there is a party obtains.

Let us now consider the essence of a party or the definition of the concept of a party. I do not think it is possible to define what a party is. Reveling is supposed to be one essence of a party in the sense that there are many parties where people are reveling. Nevertheless, there are parties without people reveling, such as a dull geek party. You can imagine that nobody in such a party revel, and they just talk about the fundamental structure of AGI, and different algorithms, as they normally do in their offices. That is to say, reveling is not an essence of party-hood. Note that here I do not deny the grounding relationship that (4) states. What I doubt is whether Fine's essentialist model can explain this grounding relationship.

One may propose that the concept of a party can be defined by a disjunction, so the essence is also disjunctive. For example, a party is either dull-style or

reveling or **whatever**. I doubt this proposal. First, without specifying all of its disjuncts and their interrelationship, it is hardly a satisfactory definition. The term “whatever” in the definiens glosses over the inadequacy of the proposal. Second, more essentially, many concepts by themselves are vague like the concept of party-hood. There are diverse parties where people act diversely on those parties. It is in principle impossible to exhaust the disjunctive statement. Finally, there will always be new kinds of parties that emerge which cannot be sufficiently characterized in existent party concepts. Nevertheless, we call them parties because they interrelate with traditional parties in a quite loose way. For many concepts, like party, or Wittgenstein’s ‘game’, they are not closed by a concept frontier.

So far, I think it is clear that we cannot give precise definitions to many concepts, or cannot specify the essence of many objects or facts. If this is the case, then Fine’s essentialist model of grounding fails in domains where precise definitions are impossible. I also do not think that this model is worth pursuing if its application is quite limited.

4. Grounding without essence

I am sympathetic to the concept of grounding. It is an explanatory powerful concept. However, I do not believe that appealing to the essence of the grounded fact is the appropriate way to explicate the grounding relationship, as argued in the preceding section. If this is the case, then how can we make sense of the grounding relationship manifested in (4)? Let us reconsider our initial intuition regarding grounding. We conceive of it as a constitutive, atemporal, dependent, metaphysical explanation. This intuition does not necessitate the involvement of essence in the explanation. That is to say, it is possible that some contingent properties or relations can fulfil the same function. I propose a pluralistic explanation; in addition to essences, recognition-related contingent properties can also ground the explanation of the grounding relationship.

We know in what context it is appropriate to characterize an event as a party, and we also know in virtue of which facts it is a party. The reason varies. For instance, Alice believes they are partying because they are reveling. Reveling is a recognition-related property that renders the event a party. A geek believes

he is partying because he and his friends are sitting together and having boring conversations. Conversing in a certain circumstance is the recognition-related property which enables another explanation for a grounding statement. The feature of recognition is essential because, in a specific case, the recognized properties enable the understanding of which subject is engaged. Still, considering the party example, why do we call it a party? Because we recognize it as a party through reveling or other recognition-related significant properties. It is crucial to note that grounding is a form of explanation, and an explanation, as ordinarily understood, requires comprehension. So positing a recognition-related property as the explanation of the grounding relationship is quite natural.

Why are philosophers driven to pursue the essentialist model? I suspect that this is because grounding is closely related to the constitutive relation, and a constituent tends to be regarded as essential to the constituted. However, this is not necessarily the case. For example, in (2), the blueness of the sky is constituted by cerulean, but the sky's blueness is contingently cerulean because, at another time, it might be pale blue. However, the contingency of the property does not undermine the grounding relationship in (2). The reason is that being cerulean is the recognition-related property in that case. Fine's generalized condition regarding any bluish color x is indeed irrelevant in explicating the grounding relationship.

In summary, grounding, as a metaphysical explanation, does not necessarily depend on the essentialist model proposed by Fine. As argued in section 3, some grounded facts have no clear-cut essence, or they simply do not have any essence, so at least for those cases, the essentialist model fails to explicate the grounding relationship. I have also argued above that to answer the question of why A is grounded in B , some recognition-related properties are required, and those properties can be contingent. More specifically, the grounding relationship holds because properties related to B are recognized as properties related to A , and such a property need not be an essence of A . It seems to me that my pluralist model can not only explicate the grounding relationships between vague facts like party-hood, but it can also accommodate simple cases involving logical and mathematical facts.

References

- Dasgupta, S. (2014) "On the Plurality of Grounds", *Philosophers' Imprint*, 14.
- Fine, K. (1994) "Essence and modality", *Philosophical Perspectives*, 8 (Logic and Language), 1–16.
- Fine, K. (2012) "Guide to Ground", in: F. Correia & B. Schnieder (Eds.), *Metaphysical Grounding*, 37-80. Cambridge University Press.
- Goff, P. (2019) "Grounding, Analysis, and Russellian Monism", in: S. Coleman (Ed.) *The Knowledge Argument*, 198–222. Cambridge University Press.
- Rosen, G. (2010) "Metaphysical Dependence: Grounding and Reduction", in: B. Hale & A. Hoffmann (Eds.), *Modality: Metaphysics, Logic, and Epistemology*, 109–135. Oxford University Press.
- Schaffer, J. (2019) "Anchoring as Grounding: On Epstein's the Ant Trap", *Philosophy and Phenomenological Research*, 99(3), 749–767.

Acknowledgement

The work is supported by The National Social Science Fund of China - Post-funding Project, 23FZX008.

Is Creeping Minimalism a Threat to the Realism-Expressivism Distinction?

Mohammadjavad Hajjalikhani (Vienna, Austria)

Abstract

Old moral expressivists believed that moral statements are not truth-apt, so they were distinguished sharply from realists. But new moral expressivists are quasi-realists: they think that moral statements are true, while expressing our pro-attitudes toward things. They argue for the truth of moral statements with minimalism about truth. If this minimalist attitude is generalized to properties and facts, expressivists can say that moral statements are factual and there are moral properties. But, then, what will be the difference between moral realism and expressivism? This is the problem of creeping minimalism. In this paper, I'll argue that if realists want to satisfactorily distinguish themselves from expressivists, they should reject minimalism. This is in contrast to the standard way of framing the creeping minimalism problem, according to which we need to state the difference between realism and expressivism in a way that is consistent with minimalism. Then I'll respond to two objections. In the end, I argue that realism is not motivated if it is stated in a minimalistic framework. So, realists are motivated to reject minimalism and successfully respond to the creeping minimalism challenge. The upshot is that there is an interesting connection between a broad kind of minimalism and moral realism.

1. Creeping Minimalism

We always say that something is right or wrong, moral or immoral; but it is not clear what we do when we make these statements (Also when we have a moral belief. But I focus on moral statements and assertions for convenience). Moral realists think that moral statements describe reality and can be true or false, as scientific or ordinary statements like “tables exist” are truth-apt and factual. Expressivists, on the contrary, think that moral statements express our attitudes toward actions: saying “helping others is good” is just saying “Hooray to helping others!”. So moral statements are not truth-apt, in the same way that you cannot say whether “goodbye!” is true or false. In short, realists and expressivists disagree on whether moral statements are truth-apt or factual.

But this way of distinguishing realism and expressivism breaks down if expressivists can accommodate ordinary moral talk—which means accepting that many moral statements are true and factual. They can do this via minimalism about truth. According to minimalism, truth “doesn’t have some deep or hidden nature. The concept of truth is not substantive and explanatory, and we use it merely for convenience: to quantify over sentences

(or propositions) and refer to them. Minimalists think that saying “p is true” just is to assert p, and this is all that can be said about truth. Following minimalists, expressivists can say “Of course ‘killing is wrong’ is true, because killing is wrong!” Our minimalist expressivists now think that moral statements are truth-apt and true, just as realists do. (See Blackburn (1993) for an implementation of this strategy.)

Expressivists can even accept that moral statements are factual since it is easy to extend their minimalism to facts: p is factual just in case it states a true sentence. So “killing is wrong” designates a fact because it is true. The minimalist expressivist believes that moral statements are true and factual, just as the realist. So, if not truth-aptness and factuality, what is the difference between realism and expressivism? One might say “Only realists believe that moral properties exist.” But expressivists can apply minimalism to predicates: if predicate P is true of something, then there is a property that P refers to. “Wrongness” is true of killing, so it refers to a moral property, wrongness. If expressivists continue to apply minimalism to all semantic notions, like reference, representation, or description, it will be hard to distinguish them from realists.

This is the problem of creeping minimalism, of framing the difference between realists and expressivists who are minimalists about all semantic notions; let’s call them broad minimalists. (My formulation of creeping minimalism is based on Dreier (2004).) The challenge is to find a thesis T that realists accept while expressivists deny—T is what realism consists in. In what follows, I argue that we can solve the problem only if the thesis T is of a special kind.

2. Solution

We can respond to creeping minimalism in two general ways. First, the way I believe is unhelpful, is trying to frame the difference—thesis T—in a way that is consistent with minimalism. For instance, T1 can be “there is a causal explanation for how we refer to moral properties”—only realists believe in a causal link between our moral talk and moral properties. Or, following Dreier (2004), T2 might be “moral properties appear in the explanation of how moral

sentences obtain their meaning.” Realists explain the meaning of moral statements by invoking moral properties, while expressivists explain it by our desires or other non-cognitive attitudes.

This is not a satisfactory response, since almost all ways of stating T leave some brand of realism or anti-realism behind. As Golub (2021: 975) noted, causal tracking responses cannot accommodate non-naturalist realists who believe in a special realm of moral properties that is not causally linked to the natural world. (See Price (2013: 36) for a conception of tracking explanations.) And, following Chrisman (2008: 348), T2 cannot accommodate non-expressivists, like Mackie, who think that moral sentences are all false, which means there are no moral properties to explain the meaning of these statements. I conclude that it is hard to state the difference in a minimalistic framework.

So let’s explore the second kind of response, which is framing a thesis T that is inconsistent with broad minimalism. In this way, realists about moral statements are committed to a non-minimalistic concept, which means they will reject broad minimalism.

Here is one possible precisification of this response. We can say that realists are committed to a non-minimalistic notion of truth or representation, and realism consists in believing that moral statements are true or representational in this sense, while expressivism denies this. We should be careful about what this “denial” means. As we’ve seen before, the minimalist expressivist doesn’t think that moral statements are false or non-representational. They are happy to say that “killing is wrong is true”, just like realists. But this agreement is superficial, since “true” has different meanings in the mouths of minimalist expressivists and realists: expressivists, unlike realists, have a minimalistic notion of truth in mind. Accordingly, we can restate T more accurately as “moral statements are TRUE”, where “TRUE” denotes the non-minimalistic conception of truth.

What do expressivists think about T? They might accept that TRUTH is a sensible concept and simply believe the negation of T. Minimalist expressivists are more likely to reject the non-minimalistic notions altogether, thinking that

they are non-sensical or not apt for philosophical theorizing. Whatever way they choose, expressivists think that there is a problem with T, and therein lies their disagreement with realists.

This way of responding to the problem of creeping minimalism might seem more like discarding the problem instead of “solving” it. Creeping minimalism presses us to characterize realism inside a minimalistic framework. As Golub (2021: 973) mentions, this is because even minimalist expressivists might want to be realists about other domains. For instance, they want to be realists about statements like “there is a table”. In other words, they still want to make a distinction between moral and ordinary objects discourses. If realism by definition contradicts broad minimalism, then minimalist expressivists cannot be realists about ordinary object discourse.

I simply admit that minimalist expressivists cannot be realists about ordinary objects, in the sense of “realism” defined above. But that is not the only brand of realism. Minimalist expressivists about morality can be minimalist realists about ordinary objects, by saying that there are causal tracking explanations for statements about ordinary objects, while there is no such explanation for moral statements (Gibbard (2015: 213) is an expressivist who employs this idea.). So “realism”, just like truth, has both minimalistic and non-minimalistic conceptions, which are two different philosophical theses. The upshot is that you cannot define realism in a way that is neutral about broad minimalism.

This connection between realism and minimalism might seem surprising and controversial. I cannot argue for it conclusively, but many realist approaches to metaphysics and morality are non-minimalistic (See, for instance, Sider (2012), Enoch (2011) and Enoch & Weinshtock Saadon (2023)). In the second objection below, I will discuss why realists are motivated to reject minimalism.

3. Objections and Replies

Objection: Tracking explanations can give us a criterion for minimalistic realism about ordinary objects. But what if our minimalist wants to be a genuine realist about a domain of abstract objects while remaining

expressivists about morality? Numbers do not participate in causal relations, so there are no causal tracking explanations of numbers discourse. Minimalist realism cannot be extended to numbers.

Reply: The minimalist you describe wants to be an expressivist about morality and a realist about mathematics. She is a minimalist about all semantic notions, including truth and singular-termhood, so she already accepts the truth of mathematical statements and the existence of mathematical objects. But she wants more than this—something to distinguish her attitude toward mathematics from her minimalist moral expressivism. She wants to be a genuine realist about numbers.

But it is confusing why our minimalist wants something more than the existence of mathematical objects. First, the standard version of mathematical realism (platonism) is accepting the existence of mathematical objects and the truth of mathematical statements, not something more than that. Consequently, it is not the case that the minimalist just wants to adopt a well-known philosophical thesis. Secondly, her genuine realism about numbers is unmotivated. Platonism about mathematics is motivated by its consistency with common sense, physical sciences, and mathematics itself: we need mathematical objects to account for the truth of arithmetic or sentences like “the number of planets is eight” (See Linnebo (2009: Part 2) for more details on the motivation behind platonism). This consistency requires only the truth of mathematical statements and the existence of mathematical entities—something which our minimalist, without her new full-blooded realism, is already committed to. What our minimalist wants to add to platonism, whatever it is, is not well-motivated.

Objection: Your definition of realism centers on how non-naturalist moral realists interpret moral claims. However, according to naturalist moral realists, moral properties are identical to (or reducible to) natural properties. Naturalist realism doesn't seem to contradict minimalism. Just as minimalists can accept the property of being a table is a natural property, they can accept that moral properties are natural properties. This casts doubts on whether your definition of realism is inclusive enough.

Reply: I agree that naturalist moral realists can accept minimalism. In this way, their realism is similar to minimalist realism toward ordinary objects: there are causal tracking explanations of our moral discourse. This means these realists are not included in our definition of non-minimalist realism. Is this a bad result?

Maybe it is bad that our definition doesn't cover one group of moral realists. It should be noted, however, that many naturalist moral realists are not motivated to accept minimalism. This is because one motivation behind their being realists (and not expressivists) is that they regard moral facts or properties as substantive or distinguished. And it can be argued that naturalist realism, without any other non-minimalistic assumption, cannot provide substantivity.

Let's first see what substantivity is. In short, substantivity is the idea that there is something deep, human-independent, and objective about morality. Let's see an example (the original example is from Horgan and Timmons (1991), but here it is used for a different purpose). Consider a moral twin earth, where there is a society just like us, with the only difference that they apply normative words "right" and "wrong" in a radically different way. Maybe they say that killing innocents for fun is "right" and have a positive attitude towards it, while helping others is "wrong". We are wrong in their opinion, and they are wrong in ours; everything seems symmetric so far. But this relativistic symmetry bothers many: there is a problem with their moral evaluations, and not us, after all! And one motivation behind realism is breaking this symmetry: taking our practice of moral evaluation to be somehow better, or distinguished, from people on the other planet. This is what we can call the substantivity of (our) moral facts.

Naturalist realism, without any additional assumption, cannot break the symmetry. I want to suggest that if you don't reject broad minimalism, you cannot provide substantivity. Let's first distinguish between semantic notions and their worldly counterparts. For every semantic notion, like predicate or singular term or sentence, there is a worldly counterpart, like property or object or fact. These counterparts can be called ideologies. Broad minimalism is the thesis that semantic notions are satisfied with worldly ideologies rather easily: if you have a true sentence with a predicate in it, you have a property

corresponding to that. Applying minimalism to our moral twin earth case, we can see that the other society's moral jargon—moral truth, moral predicates—is satisfied easily with ideologies. They are just talking about a different thing, which we can call morality*. But there is no problem with moral* properties or moral* facts: there is no ideology that constitutes our moral practice being distinguished. And apart from an ideology, it is hard to see what can constitute our difference from them (See Dunaway and McPherson (2016) for a response to this worry using a non-minimalistic notion of reference.)

4. Conclusion

I discussed how we can solve the problem of creeping minimalism by rejecting to frame the realism-expressivism distinction in a minimalistic framework. The result of this solution is that moral realists cannot be minimalists anymore. One reason for supporting this result comes from the discussion around substantivity and moral twin earth. Realists need some explanatory concept like truth or propertyhood so that they can say to their moral counterparts “Your moral statements are not true. (Or it doesn't refer to genuine properties.) The fact that you can assert “Killing is right” doesn't mean that it is true because truth needs more than assertability.” The ability to make this kind of response and thus take a serious attitude toward our moral evaluations is, I think, the core of realism.

References

- Blackburn, Simon (1993) *Essays in quasi-realism*. New York: Oxford University Press.
- Chrisman, Matthew (2008) "Expressivism, Inferentialism, and Saving the Debate", *Philosophy and Phenomenological Research* 77 (2):334-358.
- Dreier, James (2004) "Meta-ethics and the problem of creeping minimalism", *Philosophical Perspectives* 18 (1), 23–44.
- Dunaway, Billy & McPherson, Tristram (2016) "Reference Magnetism as a Solution to the Moral Twin Earth Problem", *Ergo: An Open Access Journal of Philosophy* 3.
- Enoch, David (2011) *Taking Morality Seriously: A Defense of Robust Realism*. Oxford, GB: Oxford University Press UK.
- Enoch, David & Weinshtock Saadon, Itamar (2023) "Oh, All the Wrongs I Could Have Performed! Or: Why Care about Morality, Robustly Realistically Understood", in: Paul

Bloomfield & David Copp (eds.), *Oxford Handbook of Moral Realism*. New York: Oxford University Press. 434-462.

Gibbard, Allan (2015) "Global Expressivism and the Truth in Representation", in: Steven Gross, Nicholas Tebben, and Michael Williams (eds), *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism*. Oxford.

Golub, Camil (2021) "Representation, Deflationism, and the Question of Realism", *Ergo: An Open Access Journal of Philosophy* 7.

Horgan, Terence & Timmons, Mark (1991) "New Wave Moral Realism Meets Moral Twin Earth", *Journal of Philosophical Research* 16:447-465.

Linnebo, Øystein (2009) "Platonism in the Philosophy of Mathematics", in: Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

Price, Huw (2013) *Expressivism, Pragmatism and Representationalism*. Burlington, VT: Cambridge University Press. Edited by Simon Blackburn, Robert Brandom, Paul Horwich & Michael Williams

Sider, Theodore (2011) *Writing the Book of the World*, Oxford, England: Oxford University Press.

Holistic Understanding as the Primary Epistemic Aim of Metaphysics

Jani Hakkarainen (Tampere, Finland)

Abstract

In recent years, the epistemology of understanding has begun to be integrated into metaphilosophical discussions about the epistemic aims, value, and progress of philosophy. The discourse is currently expanding to similar topics concerning metaphysics as a sub-discipline of philosophy. In this regard, I argue that holistic understanding about metaphysical matters construed from a formal ontological point of view is the primary epistemic aim of metaphysics as a field of inquiry. It makes better sense of the actual nature and practices of metaphysical study, when considered collectively, compared to the notion that metaphysics primarily seeks to provide true answers to metaphysical questions or knowledge thereof. The formal ontological approach emphasises the structured nature of ontological categories, comprising formal ontological relations, as central to the subject matter of metaphysics. Holistic understanding is an epistemic achievement precisely about a system, structure, subject matter, or body of information whose constituent parts or elements are appropriately related to one another. My proposal opens the avenue for progress in metaphysics through the acquisition of holistic understanding even though it may not offer a definitive true theory or achieve collective convergence on such a theory. I propose that if progress can be achieved in acquiring holistic understanding within the domain of metaphysics, then metaphysics can rightfully be classified as a science in the broadest sense.

1. Introduction

In times of ecological crises, our relation to nature is a hot issue. This relation is also a metaphysical matter since it concerns the relation of our existence to other beings on the Earth. Our existence is distinct from but in many ways dependent on them. Nevertheless, what constitutes a tenable metaphysical account underpinning our durable relationship with nature? We need right metaphysics of it, but how are we to decide which metaphysical view is the tenable one here? This situation underscores the need for a comprehensive and satisfying account of the epistemology of metaphysics, one that incorporates and builds upon recent advancements in the fields of epistemology, metaphilosophy and metametaphysics.

In this millennium, epistemology has seen a burgeoning interest in, among other things, *understanding* alongside knowledge, justification, and scepticism (for an overview, see Hannon 2021). In recent years, the epistemology of understanding has begun to be integrated into metaphilosophical discussions about the *epistemic aims*, *value*, and *progress* of philosophy (e.g., Dellsén, Lawler & Norton 2021, Hannon & Nguyen 2022, Keren 2023). From this

perspective, Michael Hannon and James Nguyen (2022) have argued, countering for instance David Chalmers (2015), that the *primary* epistemic aim of philosophy is understanding, as opposed to Chalmers' assumption of it being true answers to philosophical questions or knowledge of such truths. The epistemic aim of a field of inquiry is, roughly, the epistemic achievement that is the collective goal of the inquirers of the field. Consequently, Hannon and Nguyen have opened up a new viewpoint on the epistemic value and potential for progress in philosophy, challenging the sceptics of these aspects (notably Brennan 2010, Dietrich 2011, Chalmers 2015).

The discourse is currently expanding to similar topics concerning metaphysics as a sub-discipline of philosophy (McSweeney 2023, Sjölin Wirling forthcoming). Also noteworthy is Helen Beebe's (2018) view, following David Lewis, that philosophy seeks an equilibrium between common sense ("Moorean facts"), science and theoretical virtues. This view aligns closely with Gideon Rosen's (2020) proposition, according to which metaphysical inquiry aims at a set of competing theories that meet certain constraints like plausibility and theoretical virtues (Sjölin Wirling forthcoming).

McSweeney's, Rosen's and Sjölin Wirling's works, as well as Jack Ritchie's very recent 2023 paper, may be construed as an indirect response the champions of radically naturalized metaphysics, such as James Ladyman and Don Ross (2007) and Tim Maudlin (2007), and one of their sharpest critics, Kerry McKenzie (2020, 2021). For the naturalists and McKenzie assume that the primary epistemic aim of metaphysics is truth and its approximation contrasting with McSweeney and Sjölin Wirling's emphasis on understanding and Ritchie's outright rejection of truth as the aim of metaphysics. However, it should be noted that the views of Sjölin Wirling, and Hannon and Nguyen, do not imply that metaphysics does not pursue truth and knowledge at all; rather, these are not its primary epistemic aims.

My proposal contributes to the forefront of this ongoing research in the epistemology of metaphysics. I argue that the primary epistemic aim of metaphysics is *holistic* understanding about metaphysical matters construed from a formal ontological point of view. The formal ontological approach emphasises the structured nature of ontological categories, comprising formal ontological relations, as central to the subject matter of metaphysics.

2. Subject Matter and Main Branches of Metaphysics

The *premise* of my proposal hinges on the unifying formal ontological account of the subject matter and main branches of metaphysics that we have defended with Markku Keinänen in *Formal Ontology* (Cambridge University Press, 2023). Building upon Barry Smith's (e.g., 1998, 2005, Arp, Smith & Spear 2015), Peter Simons' (e.g., 1998, 2009) and E.J. Lowe's (2006) work, our account posits that metaphysics is primarily divided into general metaphysics, formal ontology, and ontology. General metaphysics considers being (what it is to be), its relation to existence, reality, becoming, and their opposites such as non-being. For example, is being the same thing as existence? Questions concerning the unity or plurality of being (ontological monism and pluralism) and the possible principles of being, such as the principle of non-contradiction, are also encompassed within general metaphysics. Formal ontology, on the other hand, scrutinizes the categories of being by analysing them through the ways in which entities are there, that is, ontological forms (e.g., existential ontological dependencies and mereological relations). For example, substances are entities existing independently, numerically identically, persistently, and in a property-bearing way in some theories. These ontological forms are formal ontological relations. Ontology deals with questions about what entities are there and why, focusing on entities and their possible grounds. It is then the task of ontologists to inquire whether entities are in grounding or fundamentality hierarchies or not and which those hierarchies are.

In the *logical* order of these main branches, ontology is positioned below formal ontology, which in turn is subordinate to general metaphysics. This means that ontological questions are framed within the context of formal ontology and general metaphysics. We inquire whether entities of a particular ontological category, such as universals, exist, or which ontological category's entities serve as the grounds of others and perhaps as the absolutely fundamental entities (like the category of substances traditionally or relations more recently). By contrast, we can analyse conceivable ontological categories by ontological forms without ontologically committing to the entities of these categories and forms. Engagement in formal ontology is feasible without adhering to specific ontological commitments, whereas the study of ontology inherently presupposes certain formal ontological assumptions. At the same

time, we inevitably make assumptions in response to the general metaphysical question of what it is to be and how being is related to existence, for instance. Ontological and formal ontological problem formulations presuppose certain aspects of general metaphysics.

Beyond these main branches of metaphysics, there are special metaphysics, such as social ontology and metaphysics of science. Each of these special metaphysics is characterized by its distinctively narrower domains of investigation. These domains are more circumscribed compared to the unrestricted domain of being that is the focus of the main branches. Our account sets special metaphysics and ontology apart from the other fields that study what there is, such as special sciences like physics and chemistry. Special metaphysics and ontology uniquely approach entities with problem settings shaped by general metaphysics and formal ontology, which impart fundamental characteristics to their specific problems. Although their domains are more narrowly defined, special metaphysics are inherently linked to general metaphysics and formal ontology. In contrast, while special scientific inquiries into entities also intersect with these philosophical disciplines, they do not centrally focus on them. For example, in exploring the existence of dark matter, the consideration of whether being is unitary or not is not a primary concern.

My proposal combines this unifying account of the subject matter and branches of metaphysics with recent literature in the epistemology of understanding and metaphilosophy. This approach will enable me to contribute not only to this literature but also to the ongoing debates regarding the scientific status of metaphysics. All of this takes place within the framework formulated in *Formal Ontology*, thereby shedding new light on the epistemology of metaphysics.

3. Holistic Understanding as the Primary Epistemic Aim of Metaphysics

Holistic or objectual *understanding* is an epistemic achievement essentially about a system, structure, subject matter, or body of information (e.g., evolution or theory of evolution) whose parts or elements fit together properly, that is, are related in a certain manner (for an overview, see Hannon 2021). Most likely, it involves *grasping* these relations and their system, rather

than being an isolated or episodic individual proposition as it can be in the case of propositional knowledge (ibid.).

This befits very well with our formal ontological account above: general metaphysics, formal ontology, and ontology, each individually and collectively, constitute a subject matter. This subject matter is comprised of various elements that share specific interrelationships. For instance, ontological categories may form a subject matter that includes, besides the categories themselves, the formal ontological relations between these categories or their entities, which are the ontological forms of the entities. They constitute a system in which some of the categories and formal ontological relations may be fundamental. Equally, ontologists strive to discover which of these categories are occupied and what the ontological hierarchies of entities are (e.g., grounding or fundamentality hierarchies). An epistemic achievement directed towards such a subject matter is referred to in epistemology as holistic understanding, as was seen just above.

This consideration, however, pertains mainly to the aspect of fittingness. A convincing argument is required to the result that holistic understanding is the primary epistemic aim of metaphysics as a sub-discipline of philosophy. A persuasive line of reasoning regarding philosophy in general appears to be present in the 2022 paper by Hannon and Nguyen. They argue that a comprehensive consideration of philosophical practices and the nature of philosophy as a field of inquiry *collectively* supports the claim that holistic understanding, rather than the formulation of a true theory on a philosophical subject or knowledge of such a theory's truth, is the primary epistemic aim of philosophy.

I propose that this argument is equally applicable to metaphysics even though it is not something that is unique to metaphysics (it is our formal ontological account of the subject matter of metaphysics that serves to differentiate it from other fields). First, in metaphysics, systematic understanding is preferred over episodic knowledge. Second, the demonstration of specific cognitive skills, such as proficient argumentation, is prioritized over mere beliefs or knowledge of the true answers to metaphysical questions in contexts of expertise, recruitment, and student evaluation practices. Demonstrating such skills is indicative of the possession of understanding. Therefore, one cannot

do good metaphysics by testimony or deference alone. Third, the field of metaphysics is characterized by enduring problems and debates, such as those concerning ontological categories (for instance, substance theories versus bundle theories), rather than a convergence on a singularly true view. Consequently, a diverse range of metaphysical theories are expected to be taught and critically surveyed in the literature. Various perspectives on a metaphysical subject matter contribute to our understanding of it. We can learn even from false historical views, such as Descartes' dualist substance-mode metaphysics. It can shed light on the possible category of modes even though it involves a false account of matter. Finally, the bottoming-out of disagreements between competing metaphysical theories, such as the disagreement between substance and process metaphysics, is widely considered epistemically valuable since in this kind of situations we understand where the disagreement lies exactly. The proposition that understanding is the primary epistemic aim makes more coherent and integrative sense of this actual nature of metaphysics as a collective cognitive endeavour and its practices compared to the assumption that its primary epistemic goal is the pursuit of truth or knowledge.

The primacy of general metaphysics and formal ontology relative to ontology suggests that ontological *explanation* is subordinate to general metaphysics and formal ontology. For instance, a potential ontological explanation that some entities exist in virtue of other entities requires the ontological categorization of the explained and explaining entities, such as into properties and objects. In this regard, I propose that ontological *explanatory understanding* is a component of formal ontological holistic understanding, which is of primary importance.

4. Prospects for Progress in Metaphysics and Its Status as a Science

My proposal opens the avenue for the prospect that metaphysics *can* progress even though it does not provide the true theory, or even converge (collectively) on it or knowledge of it. First, it would be naïve to assume that even science makes progress only when true theories are proposed and accepted (Niiniluoto 2019). It is also undecided whether metaphysical holistic understanding is factive, meaning it must centrally consist of truths, or non-factive, allowing for

the inclusion of falsehoods, possibly even centrally and intentionally (Sjölin Wirling, forthcoming). More importantly, understanding is a matter of degree: one can progress although one does not understand fully; partial understanding or such an understanding going beyond “folk metaphysics” is also epistemic progress. Indeed, *enhanced understanding* seems to be enough for constituting progress and not only promoting it.

It is widely accepted among epistemologists that holistic understanding is epistemically valuable (Hannon 2021). Therefore, if metaphysics is progressing towards holistic understanding, it is an epistemically valuable field of inquiry. Hence, pivotal questions arise: *does the field of metaphysics enhance our understanding of its central subjects*, such as ontological categories (beyond “folk metaphysics”)?

I do not delve into this inquiry here, but preliminary indications suggest affirmative evidence of such progress. In particular, it appears that there has been a notable enhancement in *our command and expansion of the modal space* concerning metaphysical subjects since the inception of metaphysical considerations millennia ago. This advancement seemingly indicates an elevated level of understanding since the command and expansion of modal space are commonly regarded as indicators of understanding (Hannon 2021). In particular, the study of formal ontology seems to have contributed to this development. The analysis of ontological categories by ontological forms has aided in commanding the modal space (Hakkarainen & Keinänen 2023). It encompasses understanding what ontological categories could be, prior to determining the actual categorization of entities. Such analysis has led to clearer, more distinct, and precise category distinctions (ibid.). Through this type of analytical approach, we can also propose novel, precisely defined categories, and their systems, thereby expanding the modal space about ontological categories. This phenomenon of expansion has already been observed in the development of trope theory, as discussed in Keinänen & Hakkarainen (2024).

Kant famously posed the question whether metaphysics at all is possible and how metaphysics in general and as a science (in the broad sense, German *Wissenschaft*) is possible. Building upon the above-said, I propose that if progress can be achieved in acquiring holistic understanding within the

domain of metaphysics, then metaphysics can indeed be classified as a science in the broad sense. In this capacity, metaphysics would meet the general criteria of science regarding understanding: it would constitute a peer-reviewed, systematic, free, publicly accessible, and collective endeavour in the pursuit of holistic understanding. Importantly, its validity would not be contingent upon any individual subject but would remain open for scrutiny or discovery by peers (cf. Hansson 2021). As previously mentioned, metaphysics may also surpass common sense epistemically in terms of better command modal space, for instance.

References

- Arp, Robert; Smith, Barry; Spear, Andrew D. (2015) *Building Ontologies with Basic Formal Ontology*, Cambridge, MA: MIT Press.
- Beebee, Helen (2018) "Philosophical Scepticism and the Aims of Philosophy", *Proceedings of the Aristotelian Society* 118 (1), 1–24.
- Brennan, Jason (2010) "Scepticism about Philosophy", *Ratio* 23 (1), 1–16.
- Chalmers, David J. (2015) "Why Isn't There More Progress in Philosophy?", *Philosophy* 90(1), 3–31.
- Dellsén, Finnur; Lawler, Insa; Norton, James (2021) "Thinking about Progress: From Science to Philosophy", *Noûs*(online first).
- Dietrich, Eric (2011) "There Is No Progress in Philosophy", *Essays in Philosophy* 12 (2), 329–344.
- Hannon, Michael (2021) "Recent Work in the Epistemology of Understanding," *American Philosophical Quarterly* 58 (3), 269–290.
- Hannon, Michael; Nguyen, James (2022) "Understanding philosophy", *Inquiry* (online first).
- Hansson, Sven Ove (2021) "Defining pseudoscience and science", in: Massimo Pigliucci and Maarten Boudry (eds.) *Philosophy of Pseudoscience*, Chicago: The University of Chicago Press, 61-78.
- Hakkarainen, Jani; Keinänen, Markku (2023) *Formal Ontology*, Cambridge: Cambridge University Press.
- Keinänen, Markku; Hakkarainen, Jani (2024) "Trope Bundle Theories of Substance", in: Anthony R. J. Fisher and Anna-Sofia Maurin (eds.) *The Routledge Handbook of Properties*, Abingdon and New York: Routledge, 239-249.
- Keren, Aaron (2023) "Disagreement, progress, and the goal of philosophy", *Synthese* 201 (2), 1-22.

Ladyman, James; Ross, Don (2007) *Every Thing Must Go: Metaphysics Naturalized*, Oxford: Oxford University Press.

Lowe, Edward J. (2006) *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Oxford: Clarendon Press.

Maudlin, Tim (2007) *The Metaphysics within Physics*, New York: Oxford University Press.

McSweeney, Michaela (2023) "Metaphysics as Essentially Imaginative and Aiming at Understanding", *American Philosophical Quarterly* 60 (1), 83-97.

McKenzie, Kerry (2020) "A Curse on Both Houses: Naturalistic Versus A Priori Metaphysics and the Problem of Progress", *Res Philosophica* 97 (1), 1-29.

McKenzie, Kerry (2021) "Science-Guided Metaphysics", in: Ricki Bliss and James Miller (eds.) *Routledge Handbook of Metametaphysics*, Abingdon and New York: Routledge, 435-466.

Niiniluoto, Ilkka (2024), "Scientific Progress", in: Edward N. Zalta and Uri Nodelman (eds.) *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition), URL = <<https://plato.stanford.edu/archives/spr2024/entries/scientific-progress/>>.

Ritchie, Jack (2023) "On the continuity of metaphysics with science: Some scepticism and some suggestions", *Metaphilosophy* 53 (2–3), 202–220.

Rosen, Gideon (2020) "Metaphysics as a Fiction", in: Bradley Armour-Garb, and Frederick Kroon (eds.) *Fictionalism in Philosophy*, New York: Oxford University Press, 28-47.

Sjölin Wirling, Ylwa (Forthcoming) "Understanding with epistemic possibilities: the epistemic aim and value of metaphysics", *Argumenta*.

Smith, Barry (1998) "Basic Concepts of Formal Ontology", in: Nicola Guarino (ed.) *Formal Ontology in Information Systems*, Amsterdam: IOS Press, 19–28.

Smith, Barry (2005) "Against Fantology", in: Johan C. Marek and Maria E. Reicher (eds.) *Experience and Analysis*. Vienna: öbv&hpt, 153–170.

Simons, Peter (1998) "Metaphysical Systematics", *Erkenntnis* 48(2/3), 377–393.

Simons, Peter (2009) "Ontic Generation: Getting Everything from the Basics", in: Alexander Hieke and Hannes Leitgeb (eds.) *Reduction, Abstraction, Analysis. Proceedings of the 31st International Wittgenstein Symposium*. Heusenstamm bei Frankfurt: Ontos Verlag, 137–152.

A Pursuit of Wittgenstein's Usage of "Chimeras" in *Philosophical Investigations* §94

Shogo Hashimoto (Tokyo, Japan)

Abstract

In *Philosophical Investigations* §94, Ludwig Wittgenstein argues that our forms of expression "send us in pursuit of chimeras" and illustrates this with the example of "a pure intermediary between the propositional *sign* and the facts." In scholarly discourse, what is meant with the term "chimera" is often regarded as mere nonsense, associated with being "wrong," "delusive," or "rooted in dogmatism." This paper seeks to demonstrate that Wittgenstein has particular nuances in mind with this terminology for a more precise understanding of his insights. To achieve this goal, it explores his earlier usage of "chimera" in his 'Lectures on Ethics' and draws upon G. E. Moore's usage in works such as *Principia Ethica*, a source alluded to by Wittgenstein. Furthermore, it scrutinizes his characterization of the term "chimera" in his *Last Writings Volume 2*. Through these historical and textual scopes, this paper aims to illustrate that what Wittgenstein identifies as a "chimera" embodies two specific features: Firstly, it is comprised of different language-games, each involving absolute and factual uses of words respectively. Secondly, this amalgamation is entirely imaginary yet is purportedly treated as if it were factual.

1. Introduction

In Ludwig Wittgenstein's *Philosophical Investigations* ("PI"), the term "chimera" appears only once in §94, which reads as follows:

94. 'Remarkable things, propositions!' Here we already have the sublimation [in original: *Sublimierung*] of our whole account of logic. The tendency to assume a pure intermediary between the propositional *sign* and the facts. Or even to try to purify, to sublimate, the sign itself. For our forms of expression, which send us in pursuit of chimeras, prevent us in all sorts of ways from seeing that nothing extraordinary is involved. [PI 2009: §94]

The central argument in this section is relatively straightforward. Wittgenstein rejects such a philosophical notion as a pure and immaterial intermediary between the propositional *sign* and the facts, or as the purified sublimated propositional sign. As noted by Gordon Baker and Peter Hacker, Wittgenstein's use of the German word "*Sublimierung*" can be interpreted as encompassing both "subliming" and "sublimating," related to "the 'higher' (most important or profound)" and "the 'purified' (unadulterated, refined)" respectively (Baker; Hacker 2005: 202). This process of sublimation, caused by our various forms of

expression, prevents us from seeing that there is nothing extraordinary hidden in our everyday language and leads us in pursuit of these sublimated entities termed "chimeras" by Wittgenstein.

But the question arises: What exactly does Wittgenstein mean by the term "chimera"? In the literature, it is often assumed that the terms "chimeras" and "chimerical" stand for mere nonsense and are synonymous with terms such as "wrong," "illusory" (Hacker 2013: 5) or "aberrant," "delusive" (Stroll 1994: 48). Sometimes it is described as "rooted in preconceptions and dogmatism" (Baker; Hacker 2005: 191), or even lumped together with terms like "phantasms, myths, superstitions" (McGinn 2013: 24). There is also a tendency to simply label as "chimeras" or "chimerical" numerous philosophical failures tackled by Wittgenstein, even though he himself doesn't explicitly qualify them as such (cf. *ibid.*). Admittedly, there are also remarks in which he uses "chimera" to refer to "the ideal description of our visual image" (PGL 1988: 110) or "the aroma of the coffee" in a more distinguished sense (MS 162(b): 59). However, it is noteworthy that "chimera" doesn't appear as frequently as terms like "wrong" or "nonsense." For a more precise comprehension of Wittgenstein's insights, this paper aims to clarify what he intends by employing the term "chimera," drawing on other relevant remarks.

To illuminate Wittgenstein's usage of "chimera," this paper delves into several works where it appears. It examines relevant discussions in his 'Lectures on Ethics' ("LE"), G. E. Moore's works such as *Principia Ethica*, which serves as a textual starting point for LE, and then Wittgenstein's *Last Writings Volume 2* ("LWII"). Through this exploration, it is demonstrated that his usage of "chimera" exhibits two characteristic features, rather than simply being dismissed as "nonsense": Firstly, it is composed of different language-games, each respectively addressing the absolute and factual uses of words; secondly, it is an imaginary construct as a whole, yet it is falsely treated as if it were factual or a matter of fact.

2. "Chimera" in LE

Let's now look at the usage of "chimera" in LE, written around 1929. This section offers a brief overview, emphasizing its position between absolute and factual uses of words.

Wittgenstein grounds the subject matter of LE in Moore's definition from *Principia Ethica*: "Ethics is the general enquiry into what is good" (LE 1965: 4). However, in LE, Ethics expands to include the concrete usage of value-related words like "good" and "right," as well as expressions such as "This is a *good* chair" or "This is the *right* road." Wittgenstein argues that these expressions possess two distinct senses: the "trivial or relative sense" and the "ethical or absolute sense" (ibid.: 5). Initially, he asserts that these expressions are meaningful solely in their relative sense, indicating adherence to predetermined standards – goals or purposes rooted in facts. For example, stating "A is a *good* runner" implies that person A meets specific factual standards, such as running a certain distance within a certain timeframe. Since these standards are factual, expressions in this relative sense merely represent statements of facts. In contrast, when expressing judgments in the ethical or absolute sense, such as "B ought to behave better," it implies that person B must behave better *necessarily* or with *logical necessity*, regardless of factual circumstances (ibid.: 7). Wittgenstein contends that asserting "the absolute good" in this sense as a fact is nonsensical.

Wittgenstein employs the metaphorical term "chimera" to emphasize there is no such thing as the absolute good. He first introduces the concept of the "world-book" or "scientific book": It would contain the whole description of the world, encompassing all value judgments in the relative sense, which essentially translate to statements of facts, along with all true scientific propositions (ibid.: 6). Notably, this world-book would lack any judgment of the absolute good, as it pertains not to facts but to logical necessity. Wittgenstein further contends that "we cannot write a scientific book, the subject matter of which could be intrinsically sublime and above all other subject matters" (ibid.: 7). Here, the absolute good is portrayed as "intrinsically sublime," a phrase resonant with both "subliming" and "sublimating," as discussed previously in the context of PI §94. Accordingly, it is deemed more profound than other subject matters and is purified or refined, indicating its absolute validity. However, if it were to be considered as a describable state of affairs within the world-book, it would, according to Wittgenstein, be "a chimera" (ibid.). He argues that "No state of affairs has, in itself, what I would

like to call the coercive power of an absolute judge" (ibid.), implying that judgments on states of affairs can only be true or false relative to facts, not absolutely or necessarily.

In summary, LE portrays the absolute good as an entity hovering between the absolute and the factual uses of words, granted not only as necessarily valid but also as a matter of fact. Wittgenstein dismisses it as nothing more than a "chimera." However, why does he opt for this term instead of simply labeling it as "nonsense"? I contend that his choice of wording is deliberate. To explore the background of this usage, the next chapter will turn to Moore's works, examining how he also employs it in a remarkably similar manner.

3. "Chimera" in Moore's Works

Let's look into Moore's works, where the term "chimera" is used. This chapter first examines *Principia Ethica*, as alluded to in LE, before extending the analysis to his other writings. It then illustrates that what he terms a "chimera" comprises multiple familiar parts, yet it ultimately embodies an imaginary construct, despite often being regarded as factual.

In *Principia Ethica*, Moore defines Ethics as "the general enquiry into what is good" (Moore 1993: 54), a characterization echoed in LE. He then emphasizes the impossibility of defining what is good (cf. ibid.: 58), arguing that concepts like "good" or "yellow" are simple notions that cannot be defined. In contrast, he maintains that complex notions like "horse" or "chimera" are definable by enumerating their constituent parts, irrespective of whether they have been heard of or seen before.

This point can be particularly traced in Moore's detailed account of "chimera" ("chimaera"). Consider his concrete description of the term:

We can, for instance, make a man understand what a chimaera is, although he has never heard of one or seen one. You can tell him that it is an animal with a lioness's head and body, with a goat's head growing from the middle of its back, and with a snake in place of a tail. But here the object which you are describing is a complex object; it is entirely composed of parts, with which we are all perfectly familiar – a snake, a

goat, a lioness; and we know, too, the manner in which those parts are to be put together, because we know what is meant by the middle of a lioness's back, and where her tail is wont to grow. [ibid.: 59]

Moore suggests that a chimera is essentially an amalgamation of several parts from real animals. Despite never having been observed, it can be described or even defined by enumerating these familiar components. This characteristic compositional aspect, where the chimera is made up of known elements, is integral to understanding its nature.

Additionally, Moore frequently employs the term "chimera" in his other works, revealing another characteristic of it. Throughout his writings, he emphasizes that a chimera is purely imaginary and lacks real existence. For instance, in his 1898 article, 'The Nature of Judgment,' he illustrates why the statement "The chimera has three heads" is false compared to the true assertion "This rose is red" (Moore 1993: 4). He grounds the falsity in the fact that the conjunction of the concepts "three" and "heads" doesn't obtain among the existents, unlike the concepts "this" and "red." The non-real feature of the chimera gets more apparent in Moore's lectures from 1910-11 (cf. Moore 1953: 251, 344-5, 372-3). Notably, he employs the term "chimera" to dismiss what some philosophers mistakenly consider to be real. For instance, he addresses the question "Is Time real?" and negates it by asserting that no such entity as Time "is real" in the sense of "exists" or "is a fact" (Ibid.: 213-4). He draws parallels between Time and imaginary creatures like centaurs or chimeras, which lack existence in reality. This critique using the metaphor "chimeras" pertains to entities that are philosophically treated as existent despite lacking actual existence.

Thus, it can be observed that what Moore refers to as a "chimera" embodies two features: Firstly, it is composed of several parts, each of which is fully known. Secondly, as a whole, it is purely imaginary, although it may be mistakenly regarded as factual or existent. In my view, Moore considers these features when employing the term "chimera," and Wittgenstein adopts this usage. This becomes evident when examining another remark in Wittgenstein's writing, which will be addressed in the following chapter.

4. "Chimera" as "False Idealization" in LWII

In a remark in LWII, Wittgenstein employs the term "chimera" to refute the misconceived notion of "pure white" (LW 1993: 48). This chapter aims to explore his criticism of it. It focuses on how, akin to Moore's usage, Wittgenstein regards a "chimera" as comprising different language-games, each involving absolute and factual uses of words, and is mistakenly deemed as factual.

In the given remark, Wittgenstein first notes:

By 'pure white' one often means the lightest of all colours, by black the darkest; but not so by pure yellow, red, etc.

White seen through yellow wouldn't become yellowish-white, but *yellow*. And yellow seen through white – should it become whitish-yellow or white? In the first case the 'white' glass acts like colourless glass, in the second like opaque glass. [ibid.]

While Andrew Lugg might rightly argue that this is one of "less well-developed remarks on colour," the primary argument can be interpreted as rejecting the idea that there is such a thing as "pure white" or "transparent white" in Lugg's account (Lugg 2021: 169). In essence, the "white" glass doesn't act like "pure white" glass in any case, but only either like "colourless" or "opaque" glass. This remark is clearly aligned with Wittgenstein's other remarks on Georg Christoph Lichtenberg: "Lichtenberg speaks of 'pure white' and means by that the *lightest* colours" (ROC-III 1993: §201); "he constructed an ideal use from the actual one. And that is not to say a better one, but one that has been refined along certain lines and in the process something has been carried to extremes" (ROC-I 1993: §3; cf. ROC-III 1993: §35). Through these remarks, the concept of "pure white" can also be understood as "sublime," signifying something "higher" or "more profound" than the actual use of color words, and also as "refined" or "purified" in an absolute sense, as seen in LE and PI §94. Regarding this concept, Wittgenstein conclusively argues: "The 'pure' concept of colour, which one is inclined to create from our normal colour concepts, is a chimera" (LW 1993: 48).

Note that the immediately following remark in LWII elaborates on "chimera" more explicitly than the previous remarks in Wittgenstein's writings. He further claims:

Instead of "chimera" I could have said "false idealization".

Perhaps the Platonic ideas are false idealizations [In original: Falsche Idealisierungen sind vielleicht die platonischen Ideen].

If there is such a thing then, someone who idealizes falsely must talk nonsense – because he uses a mode of speaking that is valid in *one* language-game in another one where it doesn't belong. [ibid.]

This remark underscores two essential features crucial for understanding the notion of "chimera," both of which resonate with Moore's usage of the term. Let's analyze each of these features in detail.

One feature is that "chimera" can be substituted with "false idealization." This false idealization, exemplified by notions like "pure white," arises from blending two distinct usages in separate language-games. As Wittgenstein acknowledges, it is not inherently false to "*introduce* a new concept of 'pure white', e.g. for scientific purposes" (ROC-III 1993: §36), meaning to envision a language-game where "pure white" serves specific functions. However, this usage diverges from the actual language-game concerning color words, such as seeing yellow through a white glass. Consequently, it is merely a false idealization to construct the ideal use of "white" that universally applies to expressions regarding particular white or whitish objects, like "a white glass." Analogous to Moore's conception of "chimera," this notion of "pure white" can also be seen as comprising different "parts," namely, language-games, each of which is familiar, yet collectively, it remains solely an imaginary construct.

The other significant feature to consider is that the false idealizations, or "chimeras" are predicated as being "the Platonic ideas," as more precisely articulated in the original German version. This statement suggests that these false idealizations bear resemblance to Plato's conceptualizations of absolute, abstract, and eternal essences such as the good or the beautiful. However, this analogy doesn't appear to conclude the narrative. Moreover, it can be argued that Wittgenstein dismisses the idealized concept of "pure white," mistakenly

treated as if it existed or were present in particular white or whitish objects, similar to how the Platonic ideas are regarded as real or factual. Once again, this aspect aligns with Moore's usage of "chimera," which opposes the philosophical notion that, for instance, Time is deemed as if it were a matter of fact.

From Wittgenstein's elucidation in LWII, it becomes evident that his metaphor "chimera" encompasses two features that are also congruent with Moore's usage of it. Firstly, the chimera is comprised of several distinct language-games, each individually familiar to us, yet collectively, it is solely an imaginary construct. Secondly, the chimerical concept, exemplified by "pure white," despite originating from a false idealization of actual language use, is mistakenly regarded as factual, echoing the essence of Platonism. I posit that these two features also resonate within PI §94. The next chapter will undertake a deeper exploration of this aspect.

5. Recursion to PI §94

As previously noted, the usage of "chimera" can be characterized in two distinct ways: Firstly, it represents an amalgamation of different language-games, crafted through "a false idealization"; secondly, it is regarded as real or factual akin to the Platonic ideas. This chapter revisits the remark in PI §94 to examine it through the lens of these two features for a better understanding of Wittgenstein's insights.

Recall Wittgenstein's discussion of "chimera" in PI §94. He concedes that it is our forms of expressions that "send us in pursuit of chimeras", wherein a "chimera" denotes "a pure intermediary between the propositional *sign* and the facts" or even the purified sublimated sign itself. Building upon the preceding analysis, "our forms of expressions" can be interpreted as two forms of our language-games: one grounded in an absolute – "purified" or "sublimated" – usage of words, and the other rooted in their factual application. Although we are familiar with each of these forms individually, their amalgamation is nonsensical and propels us toward chasing chimeras. As implied in PI §95, in philosophical discourses dealing with "propositions" (PI

2009: §94) or "thinking" (ibid.: §95) as "sublime" entities, there is a temptation not to "stop anywhere short of the fact." Instead, this can lead us astray into viewing these entities, namely, "chimeras" as matters of fact.

In analyzing Wittgenstein's usage of "chimera" across various contexts, it appears reasonable to infer that the chimeras alluded to in PI §94 also exhibit the two aforementioned characteristics. Specifically, our "pursuit of chimeras" arises from blending various language-games pertaining to both the absolute and factual usage of words. Furthermore, the "chimera" in PI §94, for instance, exemplified as the purified and sublimated sign itself, is conceptualized as a matter of fact in a somewhat enigmatic manner. Consequently, it is essential to refrain from merely construing the term "chimera" in PI §94 as indicative of "wrong" or "rooted in dogmatism." Rather, it operates as a metaphor representing an imaginary construct composed of various components, yet treated as factual.

6. Conclusion

In PI §94, Wittgenstein posits that our forms of expression "send us to pursue chimeras," encompassing such philosophical concepts as "a pure intermediary between the propositional *sign* and the facts." Within scholarly discourse, the term "chimera" is often narrowly understood as mere nonsense, associated with being "wrong," "delusive" or "rooted in dogmatism." However, this paper demonstrates that the term possesses two distinct features, taking into account its usage by Moore and Wittgenstein in other works. Firstly, what is termed a "chimera" is constructed through the amalgamation of two different language-games, each involving an absolute and factual usage, both of which are familiar to us. Secondly, a "chimera" itself is an imaginary construct, yet falsely treated as factual. I contend that this elucidation of "chimera" not only aligns with Wittgenstein's insights, particularly in PI §94, but also aids in their better comprehension.

References

Baker, Gordon P.; Hacker, Peter M. S. (2005) *Wittgenstein: Understanding and Meaning Part II: Exegesis §§1–184*, second, extensively revised edition, Oxford: Blackwell.

Hacker, Peter M. S. (2013) *Wittgenstein: Comparisons and Context*, Oxford: Oxford University Press.

Lugg, Andrew (2021) *Wittgenstein's Remarks on Colour – A Commentary and Interpretation*, London: Anthem Press.

McGinn, Marie (2013) *The Routledge Guidebook to Wittgenstein's Philosophical Investigations*, New York: Routledge.

Moore, G. E. (1953) *Some Problems of Philosophy*, ed. H. D. Lewis, London: George Allen & Unwin Ltd.; New York: Macmillan Company.

Moore, G. E. (1993) *Principia Ethica*, revised edition, ed. Thomas Baldwin, Cambridge: Cambridge University Press.

Stroll, Avrum (1994) *Moore and Wittgenstein on Certainty*, Oxford: Oxford University Press.

Wittgenstein, Ludwig (1965) "A Lecture on Ethics", *The Philosophical Review* 74, Ithaca, 3–12. [LE 1965]

Wittgenstein, Ludwig (1993) *Remarks on Colour*, ed. G. E. M. Anscombe, transl. Linda L. McAlister and Margarete Schättle, Oxford: Basil Blackwell. [Part I and Part III are referred to as ROC-I 1993 and ROC-III 1993]

Wittgenstein, Ludwig (1993) *Last Writings on the Philosophy of Psychology, Volume 2*, eds. G. H. von Wright and Heikki Nyman, transl. C. G. Luckhardt and Maximilian A. E. Aue Oxford: Blackwell. [LW 1993]

Wittgenstein, Ludwig (1998) *Wittgenstein's Lectures on Philosophical Psychology 1946–1947*, ed. P. T. Geach. New York, USA: Harvester. [PGL 1998]

Wittgenstein, Ludwig (2000) Manuscripts from *Wittgenstein's Nachlass: The Bergen Electronic Edition*, Oxford: Oxford University Press. [MS]

Wittgenstein, Ludwig (2009) *Philosophical Investigations*, eds. P. M. S. Hacker and Joachim Schulte, transl. G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte, New York: Wiley-Blackwell. [PI 2009]

Does Grounding Make Sense? A Humanistic Approach to Ordered Ontology

Iain Haslam (London, UK)

Abstract

This paper questions how a humanistic approach to metaphysics might cohere with current trends in the subject, ultimately challenging the orthodox understanding of grounding and ordered ontology as representing objective, mind-independent features of reality. Drawing on the ideas of Wittgenstein and philosophers like Bernard Williams and A.W. Moore, I suggest that grounding discourse and the notion of an ordered ontology should instead be understood as useful tools for human sense-making.

I sketch the key features of the grounding framework and highlight a tension that arises from its purported mind-independence. I then present a humanistic approach to metaphysics, which emphasises the distinctively human perspective from which philosophical inquiries necessarily arise. From this perspective, grounding talk is valuable not because it captures objective reality, but because it facilitates our ability to make sense of the world.

I also introduce the idea of ineffable knowledge - a form of tacit understanding might guide our explanatory practices but cannot be fully captured in language or thought. This ineffable knowledge may be what leads us to find ordered conceptual frameworks more intelligible or natural.

Ultimately, I posit that adopting a humanistic approach to metaphysics can still capture some of the intuitions behind grounding and an ordered ontology, but casts doubt on their mind-independence. I hope this suggests a conception of philosophy that is more faithful to the human perspective and the context of our shared form of life.

Our civilization is characterized by the word progress. Progress is its form, it is not one of its properties that it makes progress. *Typically*, it constructs. Its *activity* is to construct a more and more complicated structure. And even clarity is only a means to this end & not an end in itself.

For me on the contrary clarity, transparency, is an end in itself

I am not interested in erecting a building but in having the foundations of possible buildings transparently before me.

So I am aiming at something different than are the scientists & my thoughts move differently than do theirs. (CV 1998: 9)

1. Introduction

1. I, like many, am attracted by the current trend in meta-metaphysics that aims to ask questions not just about what exists, but what depends on what (on ‘*how* stuff exists, not just *whether* it exists’). And I’m thus equally attracted by the Grounding framework and attention to the ‘logic’ of Ground (See esp. Fine, 2001; Schaffer, 2009; Rosen, 2010). All the same, I think the orthodox understanding of Ground as describing mind-independent structure is mistaken (see Dasgupta, 2017: 74-75 for a helpful characterisation of this). And I believe that a misunderstanding of the task of metaphysics is responsible for this mistake.

2. Sadly, a full defence of this idea is well beyond the scope of such a small paper. Therefore, my goal is not to do so, but rather to put another meta-philosophical understanding – one that is informed by Wittgenstein and his successors – on the table and see what adopting it would mean for grounding. I believe it would mean that grounding discourse, while meaningful and indispensable for philosophy, does not capture how reality is fundamentally structured independent of human minds and practices. It is a useful philosophical tool that facilitates our human ability to make sense of the world from our distinctive perspective. But I also hope to show that understanding things this way can capture the intuitive appeal of focusing on grounding while maintaining a common-sense realism. Finally, I will gesture towards an account of knowledge-of-the-world’s-structure as a kind of knowing-how-to-navigate-and-make-sense-of-the-world, rather than knowledge-that it has this structure. (Again, a full defence of these views is beyond this paper, but the ideas have been forcibly argued in great detail elsewhere, mostly in the work of Bernard Williams and, more recently, A.W. Moore (see esp. Williams, 2006, 2015; Moore, 2000, 2007, 2011, 2019, 2023a. For more on the link between Williams and Wittgenstein, see Queloz and Cueni, 2021; Queloz and Krishnan, 2023)).

2. Realism & Grounding

1. I’ll first start by outlining the boring, garden-variety realism that I subscribe to. When I produce a contentful representation, what I am doing is saying that ‘things are this way’ in an external, mind-independent world. The truth or

falsity of the representation is dependent on whether things really are this way, independently of the representation itself. Often, we produce true representations that are true only from a certain *perspective*. For example, ‘it is rainy today’ may be true from the perspective of someone in a specific place and time, but false at a different spatio-temporal location. And this perspective seems to be a matter of degree too – a more general but still perspectival representation might be something like ‘grass is green’ – a true representation for beings with human visual apparatus, but not so for beings with monochromatic vision. So we might generalise and ‘eliminate’ this perspective by describing greenness in terms of wavelengths, retinas, and the rest.

2. Much recent work in metaphysics argues that we should conceptualise this independent reality as both structured and ordered. Some entities depend on other, more fundamental entities, and these dependence relations mark the world's structure. And this work assumes the further point that metaphysics is concerned with describing the fundamental nature of reality, so it should therefore capture these mind-independent dependence relations that structure the world.

3. One candidate for capturing this structure is Ground, a metaphysically explanatory hierarchical dependence relation that orders the world into levels of fundamentality. The logic of Ground is still hotly contested, and this sketch will thus be overly simplistic. The notion I'll be using will be a fairly ‘inflated’, ‘big-G’ one, to make the point more clear (see Dasgupta, 2017 and; Wilson, 2014). Ground is meant to capture what we mean when we use phrases such as ‘depends on’, ‘holds in virtue of’ or ‘because of’, and, of course, ‘grounds’. Alongside being explanatory, Ground has several key formal features, which combine to differentiate it from causal and modal relations:

Asymmetry and irreflexivity (if A grounds B, B cannot ground A),

Non-monotonicity (grounding relations can change with added information),

Hyperintensionality (necessarily equivalent statements may not be interchangeable as grounds),

Relevance (the grounds must be relevant to what is grounded).

And, finally, on the view I'm terming 'orthodox', it is a mind-independent and objective relation (cf. Thompson, 2018; Taylor, 2022).

4. To be sure, all these features have been hotly contested. But one central tension remains between Ground's purported mind-independence and its various explanatory characteristics. Consider hyperintensionality: necessarily equivalent statements may not be interchangeable as grounds. For instance, we may say 'Socrates exists' grounds 'The singleton {Socrates} exists', but not the reverse - despite their necessary equivalence. If Ground is truly mind-independent, existing in the world itself, it seems at least *prima facie* plausible that the grounding relation should hold regardless of how the relata are represented. Indeed, hyperintensionality usually arises from cognitive limitations of subjects. This tempts the idea that grounding propositions are correct only relative to some subject (e.g., perhaps depending on whether her understanding is increased). But this clearly conflicts with Ground's supposed objectivity.

5. (Recent work in irrealism about grounding allows for the notion that the relation is dependent on subjects in some way, but leaves the structured ontology untouched, so to speak. For example, Naomi Thompson floats the idea of fictionalism about grounding: whilst the grounding relation might not exist in the world *itself*, other dependence relations do and engaging in the fiction of 'grounding-talk' is useful in metaphysics in order to capture this structure (See Thompson, 2016, 2018, 2019, 2022, 2023). My own view coheres in some ways with this, but not in others, in ways I hope to make clear).

6. Thus, the prevailing orthodoxy is to conceptualise the world in terms of the ordered ontology and philosophise in a way that best captures this structure. For many, Ground is this best way. And it is the best way because it is faithful to how reality really is structured, independently of any subject-relative factors.

7. However, this points towards another tension: when metaphysicists talk about the ordered world and how to describe it, they often employ terms like "useful," "informative," and "clear." The use of these perspective-laden terms indicates that metaphysics concerns not just objective description, but also "deals with aims, interests, and values, and thus with that which is not itself

part of the basic structure of the world” and, indeed, “trafficking as it does in the notions of what is ‘particularly good’, ‘better’, or ‘best’, not only deals with aims, interests, and values, but betrays its own” (Moore, 2023a: 131).

8. The tensions between what the orthodox view aims to do (describe fundamental reality) and the tools and methods it employs (which seem to be better than others only relative to human concerns and values) are problematic for the orthodox approach, as subjectivity is anathema to its goals. As Sider puts it, “Knee-jerk realism further requires that the betterness be objective” (2012: 19). The orthodox view seeks to capture the mind-independent structure of reality, but the perspective-laden nature of its discourse and the seeming subject-relative correctness of its grounding claims suggest that it may not be fully succeeding in this aim.

3. A Humanistic Metaphysics

1. Rather than fully argue against the orthodox view here (I don't want to imply that orthodox metaphysicists can't or haven't addressed these points), I want to propose an alternative conception that I think dissolves these tensions.

2. On my own understanding of the subject, metaphysics is concerned with the making of sense. Other disciplines, like physics, strive to make sense of the world by describing it in absolute terms. But philosophy is not physics. And so we should not expect the natural-scientific method of sense making, which scythes off elements of perspective as it does, to be a model for philosophy, which is an altogether quite different discipline. Metaphysics should not attempt to make sense of the world *tout court*, but rather of what might be called *our* world. It is “an attempt, by humans, from their unique position in the world, to make sense both of themselves and of that position.” (Moore, 2023, p. 108). It is, in Bernard Williams’ famous words, *humanistic* (Williams, 2000). And the upshot of this is that philosophy – and metaphysics by extension – has “a fundamental concern with human beings and with what it takes to be one and is properly pursued, at the deepest level, from a human point of view.” (Moore, 2023: 108).

3. If this is the case, then when we ask what the best way to conduct metaphysics is, the reasons we cite (clarity, informativeness, usefulness) betray our human aims, interests, and values. As Moore notes:

“If these metaphysical deliberations of ours—about what is required of us if we are to give an optimal description of the world—are as steeped in evaluation as this (something that the very use of the word ‘optimal’ ought already to have suggested), then it seems to be legitimate and important to ask, whose evaluation is at stake and what difference it might have made if others had been doing the evaluating. As physicists we are interested in the behaviour of rocks and stars. As metaphysicians we are interested in the merits of being interested in the behaviour of rocks and stars. Is it not possible that the concerns and values that inform the second of these interests are every bit as subjective and parochial as the concerns and values that inform an interest in the behaviour of rock stars?” (Moore, 2023: 131).

4. From this perspective, grounding talk can still play a valuable role in metaphysics, but its value lies in its usefulness for human understanding rather than solely its correspondence to mind-independent reality. When we engage in grounding discourse, we are articulating the explanatory relationships that help us make sense of things. Propositions involving the world ‘grounds’ can be true and informative, but only *given* and *relative to* a certain, human, evaluative outlook. The same proposition uttered by an alien with a radically different outlook might not have the same truth-value.

5. In contrast, if we start with an overly scientific notion of metaphysics that attempts to speak about mind-independent reality shorn of perspective, we find the subjectivity of explanation (and metaphysics) intolerable and imagine something ‘behind’ the explanation, reifying it. But there is no such thing, no ‘groundons’ in the world (Dasgupta, 2017: 74), and so when we use ‘Ground’ in this metaphysical rather than everyday sense (See PI 2009: §§66 & 116), we utter nonsense – our representation has no content.

6. In sum, on a humanistic understanding of philosophy, perhaps we ought to depict the world as ordered *not* because thinking in this way is faithful to the

fundamental structure of reality, but *rather* because representing reality in this way is faithful to our thought. “(One might say: the inquiry must be turned around, but on the pivot of our real need.)” (PI 2009: §106).

7. So what *are* we doing when we talk about grounding? What are we doing when we debate its various formal features – its logic? It certainly *seems* like we’re talking *about* something, in the same way we do when we produce empirical truths such as ‘grass is green’. Well, here is one possible response, inspired by Wittgenstein's Philosophy of Mathematics:

when we describe the various formal features of grounding, we are not saying something about what the world is like, independently of us and our talk. Rather, we are stating a rule, we are ‘enunciating one of our rules of representation’ and gesturing at the rules of the game of explanation. In accurately depicting these formal features, we are not depicting the world but rather analysing our concepts themselves; forming new concepts, establishing new ways of making sense of things, and contributing to ‘a network of norms’. (See RFM PT VII, 1998; Moore, 2023).

4. Structure

1. I have gestured at a conception on which representations that use the expression ‘grounds’ to refer to the ontological substructure of the world are meaningless, since there is no such substructure. But an ‘orthodox’ Metaphysician may object that this doesn’t match our intuitions: the world *does* seem to be structured. Things seem to fit together into levels, starting with a fundamental level and ‘moving up’ in layers of derivability, which ground captures. By eschewing talk of this structure, I fail to capture intuitions about what the world is like.

2. In what remains, I want to gesture at one final idea, which I think enables us to satisfy this intuition. This idea is that, to use Polanyi’s phrase, “We can know more than we can tell”. And the point is that whilst we may not be able to *represent* (produce contentful representations of) the ordered structure of reality, we still might *know* it. Knowledge of the structure of the world is, in other words, ineffable.

3. It is ineffable since it has no representational content and does not correspond to any fact or state of affairs. However, it still constitutes a form of knowledge, as an 'enabling state' that allows us "to communicate with others... to learn from what they tell me, to arrive at various beliefs...and in countless other ways to reposition myself in the logical space of reasons." (Moore, 2007: 483). This knowledge can be considered a type of know-how: it is the knowledge of how to process and make use of other bits of knowledge. A paradigmatic example would be my understanding of a language. It is not knowledge that is true or false, as there is nothing in reality for it to correspond to. Instead, it can be evaluated as better or worse based on its ability to facilitate communication, understanding, and sense-making among humans. The 'goodness' of this knowledge is not a matter of being 'right' or 'correct', but rather of being useful and productive within the context of our situated, value-laden perspective. It is this ineffable background knowledge that enables us to form coherent concepts, make judgments, and engage in the shared practices that constitute our form of life. As Wittgenstein puts it:

Perhaps what is inexpressible (what I find mysterious and am not able to express) is the background against which whatever I could express has its meaning. (CV 1998: 16; see also *ibid.*:10).

It is the necessary condition for the possibility of meaningful representation and discourse (see Moore, 2000, esp. chapters 5-9).

4. Understanding the world as an ordered ontology may be precisely this kind of knowledge. It enables us to form a unified whole out of our explanations and representations, organising them into a logical, coherent order that we can communicate to others. It provides a shared way of making sense, which may be better than others because of its utility. Having this understanding enables us to 'see' the most fitting description or explanation: to find certain explanations more intelligible, certain descriptions more apt, and certain ways of carving up the world more natural or intuitive. It is the background against which our representations gain their sense and their usefulness for creatures like us. And while it cannot be captured in propositional form, it manifests itself in our ability to navigate the space of reasons – to make judgments, offer justifications, and engage in the normative practices that constitute our form of life.

5. It is well beyond the scope of this essay to consider this position in full, let alone defend it. But again, my point has not been to do either of these things, but rather to point out that many of the intuitions that lead contemporary metaphysicists to talk about grounding and structured ontologies may also be captured by the radically different meta-philosophical viewpoint I inhabit. My hope is that anyone already sympathetic to the humanistic self-understanding of philosophy might see some promising avenues for exploring both meta-metaphysics and Wittgenstein. But I also hope that it demonstrates that anyone *not* sympathetic to this conception cannot simply ignore noise from this side of the aisle. Contemporary Wittgensteinians shouldn't merely dismiss the current literature in metaphysics, but also (and perhaps more pointedly), current metaphysics still can't dismiss Wittgenstein.

5. Conclusion

1. In this paper, I have argued for a humanistic approach to metaphysics that reconceives the nature and purpose of grounding discourse and the notion of an ordered ontology. While these concepts have intuitive appeal and can serve as useful tools for philosophical inquiry, I contend that they should not be understood as representing objective, mind-independent features of reality. Instead, their value lies in their ability to facilitate our human capacity for making sense of the world and our place within it. To answer the title question, whilst 'Grounding' may not <make sense>, it does **make** sense.

2. To entertain these considerations would have important implications for the project of metaphysics. Grounding discourse and related notions can still play a valuable role in this project, not by revealing the objective joints of nature, but by helping us to better understand the distinctively human practices of sense-making in which we are always already engaged, and, indeed, to contribute to the making of that sense. In doing so, we can arrive at a conception of philosophy that is more faithful to the human perspective from which it necessarily arises - one which recognizes that the kind of understanding we seek is always situated within the context of our uniquely human shared form of life.

References

- Correia, F. & Schnieder, B. (2012) *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press.
- Dasgupta, S. (2017) "Constitutive Explanation", *Philosophical Issues*. 27 (1), 74–97.
- Fine, K. (2001) "The Question of Realism", *Philosophers' Imprint*. 11–30.
- Moore, A. W. (2007) "Is the Feeling of Unity That Kant Identifies in his Third *Critique* a Type of Inexpressible Knowledge?", *Philosophy*, 82 (3), 475–485.
- Moore, A. W. (2019) *Language, world, and limits: essays in the philosophy of language and metaphysics*. First edition. Oxford: Oxford University Press.
- Moore, A. W. (2000) *Points of View*. Oxford: Oxford University Press.
- Moore, A. W. (2011) *The Evolution of Modern Metaphysics: Making Sense of Things*. Cambridge: Cambridge University Press.
- Moore, A. W. (2023a) "Carving at the Joints", in: *The Human A Priori*. 1st edition. Oxford: Oxford University Press, 127–134.
- Moore, A. W. (2023b) "Wittgenstein's Later Philosophy of Mathematics", in: *The Human A Priori*. 1st edition. Oxford: Oxford University Press, 291–305.
- Queloz, M. & Cueni, D. (2021) "Left Wittgensteinianism", *European Journal of Philosophy*, 29 (4), 758–777.
- Queloz, M. & Krishnan, N. (2023) "Williams's Debt to Wittgenstein", SSRN Electronic Journal. [Online] Available from: <https://www.ssrn.com/abstract=4541123>.
- Rosen, G. (2010) "Metaphysical Dependence: Grounding and Reduction", in: Bob Hale & Aviv Hoffmann (eds.) *Modality*. 1st edition. Oxford: Oxford University Press
- Schaffer, J. (2009) "On What Grounds What", in: David Chalmers et al. (eds.) *Metametaphysics*. [Online]. Oxford: Oxford University Press, 347–383.
- Sider, T. (2012) *Writing the Book of the World*. Oxford: Oxford University Press.
- Taylor, E. (2022) "Backing Without Realism", *Erkenntnis*, 87 (3), 1295–1315.
- Thompson, N. (2016) "Grounding and Metaphysical Explanation", *Proceedings of the Aristotelian Society* 116 (3), 395–402.
- Thompson, N. (2023) "How to be an antirealist about metaphysical explanation", *Ratio*. rati. 12367.

Thompson, N. (2018) "Irrealism about Grounding", *Royal Institute of Philosophy Supplement*, 8223–44.

Thompson, N. (2019) "Questions and Answers: Metaphysical Explanation and the Structure of Reality", *Journal of the American Philosophical Association*. 5 (1), 98–116.

Thompson, N. (2022) "Setting the story straight: fictionalism about grounding", *Philosophical Studies* 179 (2), 343–361.

Williams, B. (2015) *Ethics and the limits of philosophy*. London New York: Routledge.

Williams, B. (2000) Philosophy as a Humanistic Discipline. *Philosophy*. [Online] 75 (4), 477–496.

Williams, B. (2006) *Philosophy As a Humanistic Discipline*. A. W. Moore (ed.). Princeton: Princeton University Press.

Wilson, J. M. (2014) "No Work for a Theory of Grounding", *Inquiry* 57 (5–6), 535–579.

Wittgenstein, Ludwig (2009) *Philosophical Investigations*, G.E.M Anscombe, G.H. von Wright and R. Rhees (eds.), revised 4th edn., Oxford: Blackwell [PI].

Wittgenstein, Ludwig (1978) *Remarks on the Foundations of Mathematics*, G.H. von Wright, R. Rhees and G.E.M. Anscombe (eds.), revised 3rd edn., Oxford: Blackwell [RFM].

Wittgenstein, Ludwig (1998) *Vermischte Bemerkungen. Eine Auswahl aus dem Nachlaß / Culture and Value. A Selection from the Posthumous Remains*. Ed. by Georg Henrik von Wright in Collaboration with Heikki Nyman. Revised Edition of the Text by Alois Pichler. Translated by Peter Winch. Revised second edition. Blackwell: Oxford.

Towards a Wittgensteinian Hinge Ethics: A Hinge-Epistemological Approach to Absolute Value in Wittgenstein

Jackson Hawkins (Provo, USA)

Abstract

Recent years have seen a remarkable growth of interest in 'hinge epistemology', a family of views inspired by Wittgenstein's remarks in *On Certainty* (OC). However, most scholarly work in hinge epistemology to date has been concerned with developing Wittgenstein's remarks in OC into a more systematic epistemology of empirical propositions. In the language of the *Tractatus*, this is to say that hinge epistemology has been mainly preoccupied with propositions that have a sense, are bipolar, and that picture a state of affairs. On the other hand, comparatively little attention has been paid to the possibility of applying a hinge-epistemological framework to a realm which, in the *Tractatus*, is said to lie outside the limits of sense, namely, that of ethics and aesthetics (which according to Wittgenstein are one (TLP 6.421)).

In this essay, I advance the view that ethics—understood in the narrow sense that Wittgenstein uses in his “Lecture on Ethics” (LE)—provides fertile ground for hinge-epistemological thinking, and that the conviction that ‘absolute value exists’ is apt to be understood as constituting one very important hinge in an ethical world-picture, in much the same way as Wittgenstein takes other certainties to constitute the hinges of an empirical world-picture. Moreover, I maintain that the notion of a hinge provides a helpful way of dealing with a certain paradox that Wittgenstein identifies in the “Lecture on Ethics”. I thus propose a synthesis of Wittgenstein's very late reflections in OC with his earlier thoughts on ethics.

1. Introduction

In recent years there has been a remarkable growth of interest in ‘hinge epistemology’, a family of views inspired by Wittgenstein's reflections in *On Certainty* (OC). This movement gives pride of place to the Wittgensteinian notion of a hinge, which for Wittgenstein is a point of optimal certainty necessitated by the workings of a certain world-picture. Though themselves ungrounded, these hinges ground our everyday beliefs (OC 204-5), and also serve to guide our rational inquiry as a sort of “scaffolding” (OC 221). Annalisa Coliva conveys the idea excellently: “Just as we use certain pictures to guide us in performing certain tasks... so hinges do retain a descriptive content but perform a normative function. For they guide us in selecting what can be evidence for what” (Coliva (2016): 9).

At present, most scholarly work in hinge epistemology has been concerned with developing Wittgenstein's remarks in OC into a more systematic epistemology of empirical propositions. In the language of the *Tractatus*, this is

to say that hinge epistemology has been mainly preoccupied with propositions that have a sense, are bipolar, and that picture a state of affairs. On the other hand, comparatively little attention has been paid to the possibility of applying a hinge-epistemological framework to a realm which, in the *Tractatus*, is said to lie outside the limits of sense, namely, that of ethics and aesthetics (which according to Wittgenstein are one (TLP 6.421)).

In this essay, I advance the view that ethics--understood in the narrow sense that Wittgenstein uses in his "Lecture on Ethics" (LE)--provides fertile ground for hinge-epistemological thinking, and that the conviction that 'absolute value exists' is apt to be understood as constituting one hinge of an ethical world-picture, in much the same way as Wittgenstein takes other certainties to constitute the hinges of an empirical world-picture. Moreover, I maintain that the notion of a hinge provides a helpful way of dealing with a certain paradox that Wittgenstein identifies in the "Lecture on Ethics". I thus propose a synthesis of Wittgenstein's very late reflections in OC with his earlier thoughts on ethics.

It is worth noting that a prominent exception to hinge epistemology's tendency to overlook the non-empirical is Duncan Pritchard's 'quasi-fideism', which he has advertised as a Wittgensteinian and hinge-epistemological approach to religious belief. At certain points, such as in the *Lectures on Religious Belief* (LRB), Wittgenstein describes religious beliefs as belonging to a non-empirical realm similar to that of ethics. He claims, for instance, that in religious matters empirical evidence "wouldn't in the slightest influence me" (LRB 56). As Anna Boncompagni has pointed out, this marks a salient difference between Pritchard's Wittgensteinian project and the religious thought the pragmatists James and Peirce, whom Boncompagni identifies as Pritchard's historical forerunners (Boncompagni (2022)). Quasi-fideism thus represents an important effort to put hinge epistemology to work beyond the limits of the empirical proposition. However, as I will note later on, Pritchard's project also differs from my own in important respects.

Before moving on, a few important points have to be made. First, on Wittgenstein's account we do not first discover foundational principles and then build a body of beliefs atop them in Cartesian fashion, but rather the hinges grounding a certain world-picture are "swallowed down" (OC 143) as

we begin to learn everyday facts and go about life within that world-picture. For instance, a child learning about dinosaurs and other historical topics will naturally come to take it for granted that the earth has been around for a long time (one of Wittgenstein's favorite examples of a hinge), without this ever having been made explicit to her. Secondly, the certainty associated with Wittgensteinian hinges is not merely naturalistic. Rather, Wittgenstein thinks that the certainty of a world-picture's hinges is necessitated by the very logic of the world-picture itself. As he writes in a widely-discussed passage: "But it isn't that the situation is like this: we just can't investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put" (OC 343). Returning to my earlier example, this is to say that a world-picture in which people are given to believe that dinosaurs existed in the distant past simply cannot stand without the conviction that the world has existed for at least that long. For Wittgenstein, to doubt a hinge is to jeopardize the entire door: "Someone who doubts the existence of the earth at that time is impugning the nature of all historical evidence" (OC 188).

With these points in view, I will move on to my account of what I will call 'hinge ethics'.

2. The Paradox of Absolute Value

In Wittgenstein's brief remarks on ethics in the *Tractatus*, we are given some indication of his peculiar understanding of the term. He writes, for instance, that "the question as to the consequences of an action must therefore be irrelevant." (TLP 6.422). Such remarks in the *Tractatus* already suggest that much of what we would normally call ethical discourse is not actually *ethics* in Wittgenstein's sense. Here, the "Lecture on Ethics" is particularly valuable for the light it sheds on what Wittgenstein does have in mind when he speaks of ethics. Over the course of the lecture, Wittgenstein claims that ethics is concerned with absolute value, while also insisting that no factual state of affairs can be absolutely valuable in this way. If such a state of affairs did exist, Wittgenstein writes, its prescriptive force would be such that the entire world would be obliged to recognize it as a matter of "logical necessity" (LE 7). This, he thinks, is an absurd "chimera" (ibid.). Nevertheless, Wittgenstein admits

that he is still inclined to attribute absolute value to certain experiences such as feeling “absolutely safe” (LE 8). According to Wittgenstein, these experiences “seem to those who have experienced them, for instance to me, to have in some sense an intrinsic, absolute value” (LE 10). He thus runs up against the “paradox” that an experience, which would indeed be a factual and describable event, could have the quality of absolute value:

But when I say they are experiences, surely, they are facts; they have taken place then and there, lasted a certain definite time and consequently are describable... I will make my point still more acute by saying ‘It is the paradox that an experience, a fact, should seem to have supernatural value’ (LE 10).

Wittgenstein’s way of meeting this paradox relies on a distinction between looking at a fact in terms of an empirical description, and looking at it as a “miracle”: “The truth is that the scientific way of looking at a fact is not the way to look at it as a miracle.” (LE 11). When looking at something as a miracle, Wittgenstein forcefully rejects the suspicion that “all it comes to is that we have not yet succeeded in finding the correct logical analysis of what we mean by our ethical and religious expression” (ibid.). In other words, Wittgenstein thinks that we do not just *happen* to lack the means of adequately describing in terms of empirical facts those experiences that “constantly tempt us to attribute a quality to them which we call absolute or ethical value” (ibid.). Instead, he insists that such experiences are indescribable by “their very essence” (ibid.). Wittgenstein thus maintains that there are certain special experiences that we just cannot help but view as miracles, which are inherently resistant to empirical description. These experiences are not ‘good’ insofar as they are instrumental to some end or correct with respect to a certain rule, but rather are Good in themselves.

Wittgenstein concludes his lecture with a remark about the legitimacy, so to speak, of ethics, calling it “a tendency in the human mind which I personally cannot help respecting deeply” (LE 12). Hence, although ethics in the strict sense is non-empirical and says nothing about the factual world, Wittgenstein affirms that it is a genuine and important facet of human life. This echoes his claim in the *Tractatus* that while the mystical or nonsensical cannot be said, it can nonetheless be shown (TLP 6.522).

What I wish to argue in what follows is that this ethical ‘tendency’ can be understood to swing on hinges, and to constitute part of an ethical world-picture. I contend, furthermore, that a particularly important ethical hinge is the conviction that ‘absolute value exists’. I also argue that the treatment of this conviction as a hinge helpfully supplements Wittgenstein’s approach to the paradox of absolute value by providing an explanation of how and why we come to look at certain experiences as being miraculous or absolutely valuable, despite absolute value lacking the status of an empirical fact.

The considerations of this section serve to highlight a salient difference between this account and quasi-fideism. Pritchard’s program is mainly concerned with establishing the epistemic parity of religious and empirical beliefs. Hence, Pritchard employs hinge-epistemological thinking to argue that both types of belief are grounded by hinges, and are therefore alike in this sense. Pritchard thus differentiates between what he calls fundamental religious convictions, which operate as hinges, and everyday religious beliefs, which swing on those hinges. In the introduction to a recent essay, Pritchard gives some examples of each: “that God exists, or that miracles can occur would be natural instances of the former, whereas beliefs about, say, the more arcane elements of religious teaching would be natural instances of the latter” (Pritchard (2018a):1). Although there has been some debate (Pritchard (2000), de Ridder (2019) Smith (2021), Vinten (2022)) over exactly how/which religious convictions are apt to constitute hinges, it seems at least clear that the main concern of quasi-fideism is religious belief as understood in a more everyday sense than that of absolute value used in LE.

3. Ethical Hinges and Absolute Value

Wittgenstein tells us that we encounter absolute value in special experiences which we cannot help but look at as miraculous. It is also important to note that there is some apparent variation in the experiences that Wittgenstein identifies as absolutely valuable. In the “Lecture on Ethics”, he names three such experiences: wondering at the existence of the world, the feeling of being absolutely safe, and feeling guilty, the last of which he explains as the feeling that “God disapproves of our conduct” (LE 10). Presumably, there are many other such experiences, which, despite their differences, are united by the

instantiation of a property that for Wittgenstein can only be called absolute value. The problem, as we have seen, is that absolute value is emphatically non-empirical. How, then, can Wittgenstein (and those who sympathize with him on this matter) justify the belief that absolute value exists in some sense? On what grounds can we claim that certain experiences must be seen as miraculous, if the property distinctive of miraculous experiences cannot be empirically described (aside from simply taking Wittgenstein at his word)?

This problem becomes less vexing if we view the conviction that absolute value exists not as an empirical belief, but rather as a *hinge*. As I have already noted, Wittgensteinian hinges are quite unlike Cartesian first principles. The former are not laid down in some foundational act, so that an edifice of belief can be built atop them, but rather come to be taken for granted as a consequence of the things one learns over the course of life in a certain world-picture. As Wittgenstein says, a child learning facts about a mountain “doesn’t learn at all that the mountain has existed for a long time... It swallows this consequence down, so to speak, together with what it learns” (OC 143). And after the child has thereby “got hold of a definite world-picture” (OC 167), the hinges that she has swallowed come to structure and guide her activities within that world-picture. With all this in mind, it is natural to think that certain hinges could just as well be swallowed down over the course of repeated encounters with absolute value experiences.

To clarify this idea, it will be useful to consider a child who, as she goes about life in a certain world-picture, is exposed not only to everyday facts and empirical data, but also to Wittgensteinian absolute value experiences. Indeed, she might face all of the experiences that Wittgenstein names, as well as any number of others that he does not mention. Clearly, she notices, there is something distinctive about these experiences that sets them apart from everyday interactions with the factual world. The child may or may not ever be told, in as many words, that this special property that she senses is called absolute value, but in order to treat the conviction that ‘absolute value exists’ as a hinge, it must be the case that the activities and practices which characterize her life naturally lead her to take this for granted. For example, she might learn that the impression of pure safety which she has often felt has a religious meaning (Wittgenstein himself describes this experience as the feeling of being safe “in the hands of God” (LE 10)). When this insight is

combined with further information about God (that he is an omnipotent, omnibenevolent law-giver, etc.) it is not difficult to see how the child will naturally swallow down the idea that her experiences of absolute safety have a sort of intrinsic value. She thus becomes convinced that ‘absolute value exists’ despite having never encountered the empirical fact of its existence. The conviction therefore becomes a hinge which will ground any future experiences of absolute value that the child might have, being thus integrated into an ethical world-picture. In fact, this can be so regardless of whether the child ever explicitly claims or recognizes that she is experiencing absolute value, since under normal circumstances hinges remain unspoken, in the “background” (OC 95) of our discourse (although most hinge epistemologists recognize that hinges *can* sometimes be cited overtly).

An important advantage of viewing the conviction that ‘absolute value exists’ as a hinge is that hinges are by nature non-empirical, and cannot be supported by adducing facts. Indeed, for Wittgenstein, a hinge “is as sure a thing for me as any grounds I could give for it” (OC 111). As we have seen, the paradox of absolute value follows from Wittgenstein’s belief that there is no empirical fact of absolute value, but that absolute value is also instantiated in certain experiences. However, if we view the conviction that ‘absolute value exists’ as a hinge rather than an empirical belief, then we are absolved of any need to give it an empirical basis, for an empirical basis just does not factor into hinge convictions. Instead, we can be content with the understanding that the certainty of absolute value’s existence was swallowed down at some point, over the course of exposure to the type of experience that Wittgenstein tells us we must view as miraculous. The belief that certain experiences are absolutely valuable is thus in large part grounded in the hinge conviction that ‘absolute value exists’, which itself needs no grounding, but rather is simply necessitated by the practices of the world-picture in which one lives and acts.

Of course, the idea that ethical hinges are acquired by being swallowed down is inspired by Wittgenstein’s final writings in OC, and therefore differs in important ways from the Schopenhauer-influenced program of the *Tractatus*, echoes of which are strongly present in LE. David Wiggins, for example, reads the lecture as suggesting that “value can only enter then through the subject--through the attitude of the subject's will towards the world.” (Wiggins (2004): 7). This is a far cry from the social and world-picture-focused account that I

have given above. That this is so, however, is not immediately alarming, given that my aims in this essay are not strictly exegetical but also philosophical. All the same, it is interesting to note that Wittgenstein's later work often includes remarks that seem to align nicely with the account I have just presented. For instance, in the *Lectures on Aesthetics* (LA), Wittgenstein is reported to have said the following:

A child generally applies a word like 'good' first to food... The word is taught as a substitute for a facial expression or a gesture. The gestures, tones of voice, etc., in this case are expressions of approval. What makes the word an interjection of approval? It is the game it appears in, not the form of the words. (LA 1.5).

In this passage, Wittgenstein asserts that the acquisition of aesthetic concepts follows participation in language games and the practices associated with them, and at least one version of the notes taken during this lecture has him mentioning ethics as well. Broadly speaking, this jives with what my account alleges.

4. Conclusion

Wittgenstein's early writings espouse an understanding of ethics as transcendental and associated with absolute or inherent value. Ethics is thus held to be nonsense which lies outside the limits of the empirical world, and yet Wittgenstein persists in the belief that it is in some sense legitimate and important. However, in the "Lecture on Ethics", Wittgenstein encounters the apparent paradox that absolute value is held to be non-empirical, and yet also seems to be instantiated in certain special experiences. In this essay, I have proposed a synthesis of Wittgenstein's very late epistemological remarks with his earlier work on ethics, arguing that the former provide a useful way of approaching the paradox of absolute value. This program, which I call 'hinge ethics', differs from the early Wittgenstein's philosophy in its account of the acquisition of ethical concepts. With respect to this former point, Wittgenstein's early philosophy tends towards Schopenhauerian existentialism, whereas I have recommended a socially-minded account heavily inspired by his later philosophy. For this reason, although my project is clearly not *exegetically* feasible, I think that it retains a powerful

philosophical appeal and represents a valuable study in the applicability of a hinge-epistemological framework to a novel problem.

References

- Boncompagni, Anna (2022) "Religious Hinges: Some Historical Precursors", *Topoi* 41 (5), 955-965.
- Coliva, Annalisa (2016) "Which Hinge Epistemology?", *International Journal for the Study of Skepticism* 6 (2-3), 79-96.
- de Ridder, Jeroen (2019) "Against Quasi-Fideism", *Faith and Philosophy* 36 (2), 223-243.
- Pritchard, Duncan (2000) "Is 'God Exists' a 'Hinge Proposition' of Religious Belief?", *International Journal for Philosophy of Religion* 47 (3), 129-140.
- Pritchard, Duncan (2018) "Quasi-Fideism and Religious Conviction", *European Journal for Philosophy of Religion* 10 (3), 51-66.
- Smith, Nicholas (2021) "How To Hang A Door: Picking Hinges for Quasi-Fideism", *European Journal for Philosophy of Religion* 13 (1), 51-82.
- Vinten, Robert (2022) "Wittgenstein, Quasi-Fideism, and Scepticism", *Topoi* 41 (5), 1-12.
- Wiggins, David (2004) "Wittgenstein on Ethics and the Riddle of life", *Philosophy* 79 (3), 363-391.
- Wittgenstein, L. (1965) "A Lecture on Ethics", *The Philosophical Review* 74 (1), 3–12.
- Wittgenstein, Ludwig (1966) *Lectures & Conversations on Aesthetics, Psychology and Religious Belief*, Oxford,: Blackwell.
- Wittgenstein, Ludwig (1969) *On Certainty* (ed. Anscombe and von Wright), San Francisco: Harper Torchbooks. Edited by G. E. M. Anscombe, G. H. von Wright & Mel Bochner.
- Wittgenstein, Ludwig (1933) *Tractatus Logico-Philosophicus* (ed. Ogden), New York: Harcourt, Brace.

The World and How We Know It: Wittgenstein's Realism and Metaphysics

David Hommen (Düsseldorf, Germany)

Abstract

It is still widely held that Wittgenstein's eventual abandonment of the classical picture theory of meaning in favor of the so-called use theory also entailed a definitive rejection of realist conceptions of the world and a reorientation of the primary tasks of philosophy: away from metaphysical inquiries into the structure of reality towards conceptual analyses of the structure of our language. Against this, it is argued that 1) Wittgenstein's mature philosophy, despite its pragmatist and conventionalist leanings, still involves, indeed presupposes, an epistemological realism, and that therefore 2) metaphysics as a philosophical discipline need by no means be relinquished from his point of view, even if its methods and outcomes must be radically reconceived within the framework of his semantic theory.

1. According to the picture theory of meaning, which Wittgenstein paradigmatically ascribes to Augustine and which he himself still advocated in the *Tractatus Logico-Philosophicus*, the meaning of a word is the object or property for which the word stands (cf. PI 2009: §1). In his later philosophy, Wittgenstein arrives at a completely different view, which became known as the use theory of meaning and is summarized in the core thesis: "the meaning of a word is its use in the language" (PI 2009: §43). In this view, understanding the meaning of a word no longer means knowing its referent, but rather a) applying the word correctly to objects in accordance with given linguistic conventions, b) employing it correctly in declarative, interrogative and imperative sentences and c) using it correctly in definitions, explanations and arguments (cf. PI 2009: §§6–8, 19, 75).

With this shift from reference to use, Wittgenstein has often been perceived as denying the worldly aspects of our concepts and as seeking to transform ontological questions about the objects that our words designate into semantic questions about these words. The things themselves seem to contribute nothing to the meaning of our words; on the contrary, they seem to exist only as projections of our conceptual schemes. Philosophical analysis therefore has nothing to do but to uncover the deep structures of our language, the rules of the use of words that we ourselves have given us. This impression is reinforced by occasional statements such as these:

We feel as if we had to *see right into* phenomena: yet our investigation is directed not towards *phenomena*, but rather [...] towards [...] the *kinds of statement* that we make about phenomena. [...] Our inquiry is therefore a grammatical one. (PI 2009: §90)

One thinks that one is tracing nature over and over again, and one is merely tracing round the frame through which we look at it. (PI 2009: §114)

Grammar tells what kind of object anything is. (PI 2009: §373)

On closer inspection, however, these statements by no means compel the interpretation that Wittgenstein, in his mature philosophy, renounces realism and with it metaphysics as the philosophical endeavor to penetrate to the basic features of reality. Wittgenstein does say that the grammar of words expresses the essence of things (cf. PI 2009: §371). However, this need not mean that essence *originates* in grammar. Far from denying the existence of a language-independent reality, the suggestion could rather be that the relationship between language and reality is not to be understood as one of straightforward representation, as the picture theory would have it, but as a connection of a more intricate kind. In such a context, Wittgenstein's warnings not to predicate of things what actually lies in the method of their representation (cf. PI 2009: §104), or not to interpret new conceptions as seeing new objects (cf. PI 2009: §401), might also be taken to mean that philosophers should not expect to be able to simply read off what the world is like from our linguistic constructions. In any case, Wittgenstein makes it clear that, for him, philosophical investigations are not only aimed at linguistic but also at worldly matters – but that investigations of language are, in a way yet to be elucidated, required for any philosophically instructive investigation of the world:

One ought to ask, not what images are [...], but how the word “imagination” is used. But that does not mean that I want to talk only about words. [...] I am only saying that this question is not to be clarified [...] by pointing; nor yet by a description of some process.” (PI 2009: §370)

2. The tendency to read Wittgenstein as a linguistic idealist (cf. Anscombe 1981) is nourished above all by the importance he attaches to the concepts of

practice and form of life in his theory of meaning. Wittgenstein considers linguistic activities to be nested in social practices, which in turn are rooted in the human form of life (cf. PI 2009: §§19, 21, 23, 202). Both notions emphasize the active moment of conceptualization and seem to reduce linguistic meaning to cultural preferences or general human needs.

Here, however, appearances are deceptive insofar as, for Wittgenstein, social practices and human forms of life are always embedded in a natural world. For him there is no question that the human form of life, from which our diverse language games arise, consists not only of our “[s]hared human behaviour” (PI 2009: §206) and “regular ways of acting” (CE: 397), but also of various “facts of living” (RPP I: §630), which comprise facts about both human and nonhuman nature: “we can find a ground for the structures of concepts among the facts of nature (psychological and physical)” (RPP I: §46; cf. RFM: VII, §1).

Wittgenstein very much insists that linguistic practices can have a wide variety of purposes, not all of which serve to describe the world (cf. PI 2009: §23). Therefore, there *may* be practices that do not have to answer to natural facts. However, Wittgenstein never goes so far as to claim that all linguistic practices are so detached from the world:

To invent a language could mean to invent an instrument for a particular purpose on the basis of the laws of nature (or consistently with them) [...]. (PG: §140)

It can indeed happen, and often does today, that a person will give up a practice after he has recognized an error on which it was based. But this happens only when calling someone's attention to his error is enough to turn him from his way of behaving. (RF: 121)

In contrast perhaps to the language games of morality, religion and culture, the language games of (everyday) science, engineering, etc. depend for their usefulness, and thus meaningfulness, on their successful application to empirical phenomena. This is precisely where these languages differ from games in the sense of pure conventions (cf. LWL: 12).

3. But Wittgenstein does not merely *hypothesize* that our (descriptive) linguistic practices are anchored in an extra-linguistic reality (cf. PI 2009: 241). According to him, the existence of such a reality is no less than a condition of the possibility of those practices (cf. OC: §617). “Not only rules,” says Wittgenstein, “but also examples are needed for establishing a practice” (OC: §139). Rules (definitions, explanations) are not enough to learn the use (i.e., the meaning) of words, because the use of these rules (*their* meaning) also needs to be learned. This is the quintessence of Wittgenstein’s discussion of the rule-following paradox. Ultimately, the correct use of words must be trained on the very things to which the words are to be applied:

How do I explain the meaning of “regular”, “uniform”, “same” to anyone? – I’ll explain these words to someone who, say, speaks only French by means of the corresponding French words. But if a person has not yet got the *concepts*, I’ll teach him to use the words by means of *examples* and by *exercises*. [...] In the course of this teaching, I’ll show him the same colours, the same lengths, the same shapes; I’ll make him find them and produce them; and so on. (PI 2009: §208)

It is true that samples do not dictate the use of a word by themselves – it depends on which use is intended for them (cf. PI 2009: §§30, 73). However, for anyone to learn which words are to be applied to an object, they must first know *what it is* that these or those words should be applied to:

Suppose I show someone various multi-coloured pictures, and say: “The colour you see in all these is called ‘yellow ochre’.” – This is an explanation that another person will come to understand by looking for, and seeing, what is common to the pictures. (PI 2009: §72)

Thus, it is also the way things are, and not just the way they are used in the linguistic practice, that is essential for learning the proper use of words. This also requires that the things on which the use of words is practiced have sufficiently stable characteristics and an overall regular behavior. “Every language-game is based on words ‘and objects’ being recognized again.” (OC: §455) If the world were a completely chaotic one – if objects and properties appeared and disappeared, clustered and scattered in inexplicable and unpredictable ways – our current language games would be undermined (cf. PI

2009: §142; LPE: 267; OC: §617). If, e.g., chairs unaccountably dis- and reappeared (cf. PI 2009: §80), houses dissolved into steam with no obvious cause (OC: §513), or things changed their colors incessantly (cf. MS 136: 51–2), we would no longer have any use for our familiar concepts of chairs, houses, or colors, as we would lack the criteria by which we could learn and teach, approve and criticize their use. If, on the other hand, different systematic types of things existed and different laws of nature prevailed, our concepts might also be quite different (cf. PI 2009: 241). For example, if the surfaces of things had areas that caused pain when touched, it could be natural for us to speak of pain spots in the same way we speak of color spots (cf. PI 2009: §312).

4. For these reasons, interpretations of Wittgenstein's use theory that locate the solution to the rule-following paradox solely in a communal agreement on the use of words fall short. First of all, agreement in the sense that the members of a linguistic community agree with each other on their use of words is just a judgment on the part of the members and thus itself the application of a conceptual rule to the linguistic situation, subject to its own criteria of application. Agreement in the sense that the members' *word uses* agree with each other, on the other hand, is a state of affairs that can only obtain if all members apply the *same* words to the *same* objects. (And once again, to agree with each other that the words and objects one applies are the same is not: to apply the same words to the same objects. There is not even a fact of agreeing on this matter as long as the criteria for when the users would agree with each other and when they would not are not independently fixed.)

The communitarian solution to the rule-following paradox gains only apparent plausibility from the fact that it is usually illustrated by language games that are not obviously world-related, such as adding numbers (cf., e.g., Kripke 1982: 91). As Wittgenstein explains, however, even with these practices, agreement cannot be reached without relying on certain natural givens:

Mathematicians don't in general quarrel over the result of a calculation. [...] Were it otherwise: if, for instance, one mathematician was convinced that a figure had altered unperceived [...] – then our concept of 'mathematical certainty' would not exist.

It is no doubt true that one could not calculate with certain sorts of paper and ink, if, that is, they were subject to certain strange alterations [...]. (PI 2009: 237)

In sum, although conceptual meaning is realized in linguistic practices and the human form of life, it is, for Wittgenstein, also conditioned by natural facts, both particular and universal: “What we have to mention in order to explain the significance [...] of a concept are often extremely general facts of nature” (PI 2009: §142).

5. So far, the realism implied in Wittgenstein's semantic theory is a fairly standard one. It confirms the existence of a world of objects, events, properties, relations, etc. that is fully independent of all language and thought. Yet, many interpreters who may follow the preceding line of reasoning to the extent that they acknowledge a *metaphysical* realism in Wittgenstein's theory will nevertheless refuse to credit it with *epistemological* realism as well. For, as they see it, the proven dependence of language on reality, although logically necessary, is far too general or indeterminate for one to be able to reason from known linguistic structures to corresponding worldly structures.

In fact, Wittgenstein's theory is far from excluding authentic, specific knowledge about the world. It does, however, demand a radically new understanding of this knowledge. For, another corollary of the rule-following paradox is that we cannot recognize the things to which we (learn to) apply our concepts in a way that is *itself* conceptually mediated, on pain of triggering an epistemic regress:

We say: “we couldn't use words at all, if we didn't recognize them and the objects they denote.” [...] But have we any sort of check on this recognition, so that we know that it is really a recognition? If we speak of recognition, we mean that we recognize something as what, in accordance with other criteria, it is. (PG: §118; cf. PG: §54)

Wittgenstein's problem here is in a way a variant of Meno's paradox. In order to be able to apply a concept to something (and thus gain conceptual knowledge about this thing), one must already have some *preconceptual*

knowledge of that thing (in order to know which concepts to apply to it). Wittgenstein calls this prior knowledge an “intransitive understanding” (PG: §37) and a cognition of “what is” (PG: §118). By this, Wittgenstein means a *direct acquaintance* with the essence of a thing, which does not consist in an inner imagination or feeling, but in the “familiarity” (PG: §115) with which someone deals with the thing in question, gets to grips with it, ‘owns’ it in a practical sense (cf. PI 2009: §398). It is a knowledge that is also expressed in the “certainty” (PI 2009: §474) with which one refuses to put one’s hand into a fire, or with which one heats water to a certain temperature in order to let it boil (cf. OC: §558) – which is not simply an attitude that we take towards the world, but also reflects our “acknowledgement” (OC: §378) of its realities and laws. One’s acquaintance with an object is finally also expressed in the words one applies to the object – not as a result of one’s application of a conceptual rule, but out of an instinctive reaction (cf. PI 2009: §323), as it were, from which a linguistic convention *may* develop if other language users join in this usage (cf. PG: §116).

6. These considerations show that, for Wittgenstein, there is an apprehension of reality below the level of conceptual articulation. Our fundamental knowledge of the world is practical, not intellectual: “the end is not certain propositions’ striking us immediately as true [...]; it is our *acting*, which lies at the bottom of the language-game” (OC: §204). In fact, this tacit layer of understanding enables all more explicit forms of knowledge – and thus always remains present and perceptible in them.

This is also the crucial sense in which grammar expresses essence. We cannot ultimately *state* the true nature of things; nevertheless, their nature is *shown* in the way we speak of them. As Wittgenstein puts it, nature has a say in the formation of our concepts – “but she makes herself audible in another way” (Z: §364).

The consequences of this for the prospects of metaphysical inquiries into the structures of reality become palpable when one realizes that philosophers are ultimately also playing a language game – albeit a very fundamental one. Philosophers, says Wittgenstein, do not argue about individual facts like

scientists or laymen (cf. BBB: 59). Rather, they are dissatisfied with our (ordinary or scientific) language as a whole, because it seems too narrow or one-sided to them. “Thus we sometimes wish for a notation which stresses a difference more strongly [...], or one which in a particular case uses more closely similar forms of expression [...]” (BBB: 59).

What philosophers strive for, in other words, is a “surveyable representation” (PI 2009: §122) of the phenomena in which certain conceptual problems disappear. How they may arrive at such a representation is an important question that must be addressed in detail elsewhere (cf. Hommen 2022). At this point it should just be noted that, even if philosophers succeed in such a representation, they can never fully *conceptualize* reality with it. All concepts and descriptions are channeled through norms and interests (cf. RFM: IV, §33; PI: §570) and therefore cannot be taken at their word (so to speak).

However, this does not mean that philosophers are forever barred from the possibility of genuine world knowledge. For a *successful* surveyable representation of the world would reveal its basic features in an unspeakable way, as it would be borne by an all-encompassing practical acquaintance with this world, which in turn would evince itself in that particular form of representation.

Philosophers therefore do not need a new method. (Conceptual analysis is fine.) They only need to be clear about the nature of their method and the results they can achieve with it. Their task is not to capture and confine reality in ever more exact terms. It is rather, in Adorno's words, “to say, through mediation and contextualization, what cannot be said *hic et nunc*” and to achieve “the utopia of cognition [...] by unlocking the non-conceptual by means of the concept” (Adorno 2008: 74).

In light of this, it becomes clear why Wittgenstein says of philosophy that one should write it “only as one writes a poem” (CV: 28). A metaphysics practiced as described above would indeed be a kind of poetics which, if done well, would put us in a mood “in which one's thoughts seem as vivid as nature itself” (CV: 75). It would be the attempt to receive a “strong and lasting impression” (PH: 183) from the grammatical facts of our language, which is to

say a quasi-mystical experience of reality, and to find the “liberating word” (PH: 165) that brings this experience to its best poetic expression.

References

Adorno, Theodor W. (2008) *Lectures on Negative Dialectics*. Translated by Rodney Livingstone. Cambridge/Malden: Polity Press.

Anscombe, G. E. M. (1981) “The Question of Linguistic Idealism”, in: *The Collected Philosophical Papers of G. E. M. Anscombe. Vol. I: From Parmenides to Wittgenstein*, Oxford: Basil Blackwell, 112–133.

Hommen, David (2022) “Wittgenstein, Ordinary Language, and Poeticity”, *Kriterion* 35(4), 313–334.

Kripke, Saul (1982) *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press.

Wittgenstein, Ludwig (1969) *Preliminary Studies for the “Philosophical Investigations”*. Generally known as *The Blue and Brown Books*. Second edition. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1974) *Philosophical Grammar*. Ed. by Rush Rhees, translated by Anthony Kenny. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1974) *On Certainty / Über Gewißheit*. Ed. by G. E. M. Anscombe and G. H. von Wright, translated by Denis Paul and G. E. M. Anscombe. Reprinted with corrections and indices. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1978) *Remarks on the Foundations of Mathematics*. Ed. by G. H. von Wright, R. Rhees and G. E. M. Anscombe, translated by G. E. M. Anscombe. Third edition, revised and reset. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1980) *Wittgenstein's Lectures: Cambridge, 1930–1932*. Ed. by Desmond Lee. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1980) *Remarks on the Philosophy of Psychology*. Vol. 1. Ed. by G. E. M. Anscombe and G. H. von Wright, translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1981) *Zettel*. Ed. by G. E. M. Anscombe and G. H. von Wright, translated by G. E. M. Anscombe. Second edition. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1993) “Notes for Lectures on ‘Private Experience’ and ‘Sense Data’”. Ed. and introduced by David G. Stern. In: *Ludwig Wittgenstein. Philosophical Occasions 1912–1951*. Ed. by James C. Klagge and Alfred Nordmann. Indianapolis and Cambridge, USA: Hackett, 200–288.

Wittgenstein, Ludwig (1993) "Philosophie / Philosophy". In: *Ludwig Wittgenstein. Philosophical Occasions 1912– 1951*. Ed. and introduced by James C. Klagge and Alfred Nordmann. Indianapolis and Cambridge, USA: Hackett, 158–199.

Wittgenstein, Ludwig (1993) "Remarks on Frazer's Golden Bough / Bemerkungen über Frazers Golden Bough". In: *Ludwig Wittgenstein. Philosophical Occasions 1912–1951*. Ed. and introduced by James C. Klagge and Alfred Nordmann. Indianapolis and Cambridge, USA: Hackett, 115–155.

Wittgenstein, Ludwig (1993) "Ursache und Wirkung: Intuitives Erfassen / Cause and Effect: Intuitive Awareness". In: *Ludwig Wittgenstein. Philosophical Occasions 1912–1951*. Ed. and introduced by James C. Klagge and Alfred Nordmann. Indianapolis and Cambridge, USA: Hackett, 370–426.

Wittgenstein, Ludwig (1998) *Vermischte Bemerkungen. Eine Auswahl aus dem Nachlaß / Culture and Value. A Selection from the Posthumous Remains*. Ed. by Georg Henrik von Wright in Collaboration with Heikki Nyman. Revised Edition of the Text by Alois Pichler. Translated by Peter Winch. Revised second edition. Blackwell: Oxford.

Wittgenstein, Ludwig (2000) *Wittgenstein's Nachlass. The Bergen Electronic Edition*. Ed. by the Wittgenstein Archive at the University of Bergen under the direction of Claus Huitfeldt. Oxford: Oxford University Press.

Wittgenstein, Ludwig (2009) *Philosophical Investigations / Philosophische Untersuchungen*. Ed. by P. M. S. Hacker and Joachim Schulte, translated by G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. New York: Wiley.

Tractatus and “Radical Translation” Superimposed

Herbert Hrachovec (Vienna, Austria)

Abstract

In a surprising statement Wittgenstein's *Tractatus Logico-Philosophicus* (TLP) claims that everyday language is “in perfect order”, despite its role in disguising thought and causing confusion in logical terms. While the TLP aims for a logically regulated language, it lacks guidance on bridging this ideal with everyday language's irregularities. Wittgenstein's later works reverse this perspective, aligning with a sentiment from his early diaries that asserts the inherent sense of propositions used by humanity without awaiting critical analysis.

The submission proposes that Donald Davidson's “radical interpretation” provides a hermeneutic scenario wherein Wittgenstein's TLP approach can to some extent be seen to anticipate his later philosophy. It argues against dismissing significant parts of the TLP, suggesting a connection between its logical rigor and Wittgenstein's revised philosophical stance. The discussion weakens key TLP theses to enhance the link between its logical rigorism and Wittgenstein's later open-minded philosophy.

Four issues raised by the TLP are explored: fixation of simple things, unresolved ontology, difficulties with the context principle, and Wittgenstein's unique approach to quantification. It is suggested that Davidson's theoretical framework for understanding completely foreign languages makes available a dialogical environment for this examination. Davidson's proposal involves segmenting actual signals emitted by indigenous populations in specific situations to confirm shared understanding, implying a concept of truth reliant on mutual dependencies and verifiable sentences within a common environment. It can serve as intermediate stop between Wittgenstein's two distinct periods.

Wittgenstein's *Tractatus logico-philosophicus* (TLP) contains quite a few surprising and mysterious remarks. Hardly any of them have received less attention than his statement “In fact, all the propositions of our everyday language, just as they stand, are in perfect order.” (TLP 5.5563) This is the same colloquial language that is said to disguise thought (TLP 4.002) and to confuse the meanings of “is” as copula, equal sign and predication of existence (TLP 3.323). Wittgenstein's goal is a logically regulated language and the TLP lacks any indication of what a bridge between this ideal and the irregularities of everyday usage might look like. As is well known, Wittgenstein reversed the perspective in his later works. He thus reverted to a motif that can already be found in his diaries on 17 June 1915: “But surely this is clear: the propositions which are the only ones that humanity uses will have sense just as they are and do not wait upon a future analysis to acquire a sense.” (NB: 62) In the TLP,

this idea seems to be out of place. The present contribution provides a hermeneutic scenario in which Wittgenstein’s TLP-approach itself can to a certain extent be regarded as a precursor of his subsequent philosophy. From this point of view, the recommendation not to take large parts of the TLP seriously (cf. Conant 1989, Diamond 1991, Read & Lavery 2011, Bronzo 2012) has fallen foul of the honey pot that Wittgenstein set up at the end of his work.

In the following contribution, central theses of the TLP are weakened in order to strengthen motives establishing a connection between its logical rigorism and the open-mindedness of Wittgenstein’s revised philosophical position. How can a mitigating the methodological cornerstones of the TLP achieve such a rapprochement? It will be shown that four issues raised by the TLP can be released from the straitjacket of the “most general propositional form” (TLP 4.5) and ported into a dialogue-oriented environment. In brief, this involves the (1) fixation of simple things, (2) their unresolved ontology, (3) difficulties concerning the context principle, (4) Wittgenstein’s idiosyncratic approach to quantification. The following section exposes and illustrates the initial premise which the ensuing section spells out by tracking its application on the topics mentioned.

1. Radical Interpretation

Following on from W.v.O. Quine’s thought experiment of decoding a jungle language (Quine 2013), Donald Davidson has proposed a theoretical framework to reconstruct the understanding of completely foreign languages (Davidson 1984). Interpreters meet members of an indigenous population who, they assume, communicate by linguistic means. How to establish that they do not emit irregular sounds? Davidson regards this as a matter of segmenting their actual signals emitted in situations that the interpreters take to be assertions. As soon as a hare appears at the roadside, they may hear the exclamation “gavagai” and presume that it refers to the animal known to the listeners. To confirm the hypothesis, repetitions and tests of the episode are necessary. Their aim is to ascertain that both sides agree on the occurrence of something called hare or gavagai respectively.

“Agreement” implies a concept of truth. Sentences with a claim to validity are often connected to each other within a network of mutual dependencies. In

order for “gavagai” to be understood as a statement of the appearance of a hare, this sequence of sounds must be repeated appropriately, it must occur when such an animal is sitting far away in the field and should not be uttered when a deer suddenly crosses the meadow. According to Davidson, a Tarski theory of truth fulfils these conditions. “At the center stands a formal theory, a theory of truth, which imposes a complex structure on sentences containing the primitive notions of truth and satisfaction. These notions are given application by the form of the theory and the nature of the evidence.” (Davidson 1984: 137) The idea behind Davidson’s proposal has proven to be extraordinarily fruitful. In order to establish understanding with initially incomprehensible speakers one must work in a shared environment supposing that they express verifiable sentences. The following illustrations serve to establish a connection between these considerations and the TLP.

Wittgenstein imagines “a white surface with irregular black spots on it”. (TLP 6.341) A descriptive grid is used to record their distribution: “I can always approximate as closely as I wish to the description of it by covering the surface with a sufficiently fine square mesh, and then saying of every square whether it is black or white.” (TLP 6.341) Here is an image of haphazard (in this case colored) patches:



A variant of the above tableau superimposes a geometric grid onto its blurred contours, emphasizing its color-clusters.



Both images lack obvious meaning. They do not seem to be *about* something. We lack a key to consider them as informative – unless we view them in their “proper” mode. Given a more refined resolution a trained eye will immediately recognize it as a section taken from a hand-drawn map.



Given this image the grid depicts the structure of a settlement adjacent to green spaces. It indicates that streets run between houses at a particular location. The drawing is organized to match states of affairs. The pictorial

elements may be understood to stand for tangible facts according to familiar conventions. The choice of an appropriate pattern has turned blurred blobs into a (kind of) statement.

Visual impulses are *recognized* as depiction by means of a suitably applicable pattern. Assuming it is part of a map the image can be seen as topographical information. The guiding idea of this kind of semantics has been aptly characterized by Ernst Tugendhat in an essay on G. Frege's concept of "Bedeutung". He introduces the term "truth-value potential" (Tugendhat: 234, cf. McGuinness 2002: 87) in order to designate the contribution sentence-parts provide to determine the truth value of those sentences as a whole. In our illustration the "truth-value contribution" of the black cross (on top of the reds spot) is for this icon to refer to a church. Images, it has to be admitted, do not by themselves carry extra-pictorial meaning, which brings us to TLP's picture theory which is meant to reflect the structure of facts. Elementary propositions, as it were, capture the world's multiplicity with the appropriate resolution. Their components each contribute in different ways to the formation of a judgeable expression. Following these preparatory steps Davidson's theoretical approach will now be tentatively applied.

2. De-radicalizing the *Tractatus*

Wittgenstein's TLP and Philosophical Investigations (PI) share an essential initial feature which becomes prominent under the lens of Davidson's proposal. Augustine remembers that as an infant he was listening to "words, which I heard uttered in their respective places in various sentences" (PI 2009: §1). And he succeeded to figure out their meaning, a clear case of radical translation. Almost all items of the Davidsonian toolset are present: utterances taken as words-in-sentences according to grammatical rules. Significantly missing are truth-value potentials. There is a story to be told about how they got lost. It will be developed with regard to the four topics listed at the beginning of this paper.

(A) Simple objects. Bare ontological statements initiating the TLP have led to extensive discussions. Take "Objects are simple." (TLP 2.02) and "Objects make up the substance of the world." (TLP 2.021) Opinions differ as to whether these theses can be confirmed empirically or whether they stem from a reflection on

the linguistic coupling of "simple" and "composite" (McGinn 2007, Tejedor 2003). Wittgenstein's statements are put forward as dogmas to which nothing can be added. In the light of "radical translation", however, they appear in a different light. The first encounter with completely unknown sequences of sounds must inevitably distinguish block-like acoustic segments in order to handle the foreign idiom. Such points of reference cannot simultaneously serve as a basis for understanding *and* be overruled. Language learning is alien to the TLP, but its insistence on simple things emphasizes a general trait: knowledge cannot arise without starting points. Segmentations may be reversible, but there is no way around segmentation as take-off positions for any further steps.

A similar consideration applies to objects posited as the substance of the world (TLP 2.021). Simple things cannot be featureless: their presence would be unnoticeable. "Gavagai" includes visual and ethological clues that are predictably repeated. Traditionally, a distinction has been made between accidental and essential properties. The former ("external relations") according to Wittgenstein do not concern logic, the latter ("internal relations") are "build into" things by his ontology (TLP 2.01231). He has a case here. Recognizing something as an enduring object seems to imply an ultimate fallback on non-negotiable instances. If "gavagai" is also voiced when toasting, cursing or in battle, we have lost hold on this expression's meaning. Wittgenstein noted this motif *mutatis mutandis* later on: "Once I have exhausted the justifications, I have reached bedrock, and my spade is turned. I am inclined to say: This is simply what I do." (PI 2009: §217) To give it a Davidsonian twist: we have arrived at common ground in understanding strange pronouncements if the principle of charity requires that we countenance its use.

(B) Realism? Wittgenstein's bold ontological pronouncements suggest a straightforward realistic outlook (TLP 2.06). But notice, on the other hand, the lack of an elaborated theory of science. He is not interested in drawing a picture of factual empirical research (Ishiguro 1989). A sketch of representation via postulated elementary propositions and their logical connections is all we got. The formal apparatus is described in detail, while its required application figures much less prominently. "Logic must look after itself." (TLP 5.473) But then: "The application of logic decides what elementary

propositions there are. What belongs to its application logic cannot anticipate.” (TLP 5.557) The question is how these remarks fit together.

Davidson does not contrast practical understanding with self-contained logic. His Tarskian truth theory is constructed by registering true sentences and tracking their consequences, given changing circumstances and further assertions. The problem of how logical forms relate to the actually existing world is thus defused. Davidson by definition begins with everyday spoken language and we can see Wittgenstein taking this very step in 1929. “How strange if logic were concerned with an ‘ideal’ language and not with ours. For what would this ideal language express? Presumably, what we now express in our ordinary language ...” (PB 1998: 52) The TLP, without further ado, asserts that all everyday language sentences are just as legitimate as logic itself. Davidson’s approach supplies a strong supporting argument: Logical forms underlying a shared understanding of given environments have to be gleaned from existing practice. Augustine was listening to everyday language if there ever was one. Perceiving, sorting out and redeploying words he entered a linguistic space simultaneously preformed and performed.

(C) The context principle. Frege assigned the same semantic category (names) to simple terms and sentences. This poses a difficulty for Wittgenstein's elementary propositions. He advocates the context principle: “... only in the nexus of a proposition does a name have meaning.” (TLP 3.3) But then again: those names are supposed to single-handedly pick out simple objects (TLP 3.203, 3.22f) *and* to be part of elementary propositions (whose meaning is itself a name). Two different types of “property” are at play here. “If I know an object ... I must know all its internal properties” (TLP 2.0123), whereas external properties are optional (TLP 4.122). Contingent qualities determine the material content of elementary propositions and yet Wittgenstein’s world is essentially unchanging. “Objects make up the substance of the world.” (TLP 2.021) Factual observations (elementary sentences) cannot change internal properties, which are pre-propositional and not subject to control by truth-value-potentials. Logic, which has to be self-contained, must include an inventory of forms that covers all possible constellations occurring in states of affairs. But how can it do this in advance? How many places are to be provided (unalterably!) for the truth-value potential of a name? The situation resembles the insufficiency of a database

supposed to register all inhabitants of a city without prior information of the required syntax of future immigrant's names.

The components of a proposition, we have noted, each fulfil specific functions in determining whether or not it is true. Davidson offers a plausible criterion to distinguish between internal and external properties. To determine essential qualities look at those truth-value potentials that cannot fail to come out true of something unless you miss the reference. External properties are contingent add-ons. Compare this to Wittgenstein's distinction between ordinary experiences of how things turn out to be and his logic's "experience" *that something is* (TLP 5.552). If conformity with logic is granted to everyday language this separation cannot be maintained.

(D) Generality. Autonomous syntax is coupled to its application which it pre-determines and traces at the same time. The TLP, at first glance, does not contain "practical" information about how science is done. The book culminates in TLP 6, which defines the general propositional form containing elementary propositions as basic input, without going into details about the their (and the world's) composition (cf. TLP 5.552). On closer inspection Wittgenstein, however, gives an account of how the encounter between logic and states of affairs may be conceived in unison. It is hidden in an astonishing directive: "I dissociate the concept all from truth-functions." (TLP 5.521) Wittgenstein does not conceptualize generality the usual way, i.e. as quantors subsuming all elements of a given set. His solution is built on the capacity of truth-value potentials.

I call any part of a proposition that characterizes its sense an expression (or a symbol)." (TLP 3.31) "An expression presupposes the forms of all propositions in which it can occur. (TLP 3.311)

Wittgenstein calls such expressions "sentence variables". "Thus an expression is presented by means of a variable whose values are the propositions that contain the expression." (TLP 3.313) Given these definitions Wittgenstein proceeds to establish sets of propositions which share constitutive components, namely expressions presenting the truth-value potentials discussed above. All statements about Cambridge are, for example, similar in that they contain sentence elements to which exactly this city's characteristics

apply. Thus defined sentences represent general features of the state of affairs depicted by "generalized" constants (TLP 3.315). Note that those variables are formal tools to basically pick out "words, which I heard uttered in their respective places in various sentences". Generality is achieved by assembling all propositions containing the expression of a particular truth-value potential. A crucial question is still unresolved, however. Those propositions do not aggregate on their own. What or who selects according to which criteria (cf. Hrachovec, forthcoming)?

3. Conclusion

We have engaged in closely reading the TLP. According to a widespread alternative this is a wasted effort. In conclusion, and by way of contrast, here is a brief rejoinder. The issues of "simple things", "essences" and "generality" have been shown to span the distance between TLP and PI. Concerning generality Wittgenstein's rejection of Frege-Russell-style quantification has not often been noticed and even less often been appreciated (but cf. Scanlan 1995, Campbell 2014). This has been an unfortunate oversight resulting in the TLP being widely read as an exercise in logical regimentation, Wittgenstein himself being among the critics (e.g. PI 2009: §217). Hidden within the TLP's general sentence form, however, sentence variables do not merely represent sets of empirically encountered records, but -- reaching out into the opposite direction -- *Urbilder* (prototypes) (TLP 3.24). Those are basic patterns shown by possible instances of elementary sentences; constitutive abstractions reminiscent of Platonic forms. It is easy to overlook that those structural determinants cannot be entirely set in stone. This is the point of their dependence on use, which we have accentuated by introducing the Davidsonian angle. To draw one additional "family resemblance" in closing: *Urbilder* are precursors of the paradigms well known from the PI. Their appearance in the TLP is somewhat enigmatic, but they can be regarded as foreboding the problem of rule-following. Early Wittgenstein's logical constructions re-emerge as topics of contemporary concern. It would be a shame to tear down the bridge.

References

- Bronzo, Silver (2012) "The Resolute Reading and Its Critics. An Introduction to the Literature", *Wittgenstein-Studien* 3 (1), 45-80. <https://doi.org/10.1515/wgst.2012.45>.
- Campbell, Christopher (2014) "Categorical Indeterminacy, Generality and Logical Form in Wittgenstein's *Tractatus*". *European Journal of Philosophy* 22, 1, 138–58. <https://doi.org/10.1111/j.1468-0378.2011.00490.x>.
- Conant, James (1989) "Must We Show What We Cannot Say?", *The Bucknell Review* 32 (1).
- Davidson, Donald (1984) *Inquiries into truth and interpretation*. Oxford: Oxford Univ. Press.
- Diamond, Cora (1991) *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. Cambridge, Mass: The MIT Press. <https://doi.org/10.7551/mitpress/5797.001.0001>.
- Hossein Khani, Ali (2020) "Davidson's Wittgenstein". *Journal for the History of Analytical Philosophy* 8, Nr. 5. <https://doi.org/10.15173/jhap.v8i5.3969>.
- Hrachovec, Herbert (forthcoming) "Against auto-da-fé. A sanguine reading of Wittgenstein's *Tractatus*" in : Alois Pichler, Esther Heinrich-Ramharter, Friedrich Stadler (eds.) *Proceedings of the 44th International Wittgenstein Symposium in Kirchberg*.
- Ishiguro, Hidé (1989) "Die Beziehung zwischen Sprache und Welt im früheren und späteren Wittgenstein". *Grazer Philosophische Studien* 33, Nr. 1: 49–66. <https://doi.org/10.1163/18756735-90000390>.
- McGinn, Marie (2007) "Simples and the Idea of Analysis in the *Tractatus*", in: Guy Kahane (ed.) *Wittgenstein and His Interpreters: Essays in Memory of Gordon Baker*. Malden, Mass: Blackwell, 2007.
- McGuinness, Brian (2002). *Approaches to Wittgenstein: Collected Papers*. London: Routledge.
- Quine, Willard Van Orman (2013) *Word and Object*. New ed. Cambridge, Mass: MIT Press.
- Read, Rupert J., Lavery, Matthew A. (eds.) (2011). *Beyond the Tractatus Wars: The New Wittgenstein Debate*. New York: Routledge.
- Scanlan, Michael (1995) "Wittgenstein, Truth-Functions, and Generality". *Journal of Philosophical Research* 20: 175–93. https://doi.org/10.5840/jpr_1995_27.
- Tejedor, Chon (2003) "Sense and Simplicity: Wittgenstein's Argument for Simple Objects", *Ratio* 16, Nr. 3, 272–89. <https://doi.org/10.1111/1467-9329.00222>.
- Tugendhat, Ernst (2017) *Philosophische Aufsätze*. Suhrkamp Taschenbuch Wissenschaft 1017. Frankfurt am Main: Suhrkamp.

Wittgenstein, Ludwig (1992) *Tractatus Logico-Philosophicus*. trans. by D.F. Pears and B. McGuinness. London; Routledge and K. Paul.

Wittgenstein, Ludwig & Wright, G. H. von. (1998) *Notebooks 1914 - 1916*. Chicago, Ill.: Univ. of Chicago Press.

Wittgenstein, Ludwig (1998) *Philosophical Remarks*. Rush Rhees (ed.), trans. Raymond Hargreaves and Roger White. Oxford: Basil Blackwell.

Wittgenstein, Ludwig, G. E. M. Anscombe, P. M. S. Hacker, und Joachim Schulte (2009) *Philosophische Untersuchungen/Philosophical investigations*. Rev. 4th ed. Chichester, West Sussex, U.K.; Malden, MA: Wiley-Blackwell.

Priority Monism and the Possibility of Gunk

Alina Sophie Jacobs (Vienna, Austria)

Abstract

In *Monism: The Priority of the Whole* (2010), Jonathan Schaffer argues for reestablishing priority monism as a doctrine in contemporary metaphysics. Priority monism holds that there is a maximal concrete entity, such as the cosmos or the world, from which all other concrete entities are proper parts, with the cosmos maintaining an asymmetrical dependence relation of priority between its parts. By reconstructing the central aspects of priority monism and arguments favoring it, this paper explores and challenges one of Schaffer's arguments: the argument from gunk. One of the main arguments favoring monism is that it can comply with the possibility of gunk worlds while pluralist opponents cannot. In mereology, a gunky object is a composite material object in which each part is said to have proper parts and lacks any ultimate parts. According to Schaffer, only monists can consistently explain the possibility of atomless gunk and integrate it into their account. Based on an in-depth engagement with Schaffer's work, this paper aims to present four objections to Schaffer's argument from gunk and to provide a more profound understanding of sparse ontology.

1. Introduction

What are the basic building blocks of reality? This question is central to many debates in contemporary metaphysics but is answered by its practitioners in several distinct ways. One popular way of thinking about the issues is asking what entities are fundamental. A pluralist may regard the cosmos as pieced together from many tiny parts, while a monist may argue that the cosmos is the only fundamental thing.

In his paper *Monism: The Priority of the Whole* (2010), Jonathan Schaffer proposes to revive the monistic tradition based on developments in physics and mereology. To support priority monism, Schaffer offers a range of arguments relevant to recent debates over conceptions of metaphysical structure, such as *the argument from commonsense*, *the argument from quantum emergence*, or *the argument from gunk* (Schaffer 2010: 42-61).

Schaffer's theory of priority monism is widely discussed in contemporary metaphysics. In particular, his argument from gunk has occupied the research literature and fuelled multiple discussions in the domains of fundamentality, mereology, and grounding (see Gibberman 2021, Trogdon 2017, or Mormann 2014). This paper aims to challenge Schaffer's argument from gunk by raising four objections to the overall argument and its premises.

My paper is structured as follows: First, I will clarify Schaffer's notion of priority monism, focusing mainly on his articles *On What Grounds What* (2009) and *Monism: The Priority of the Whole* (2010). Then, I will address one of Schaffer's arguments for priority monism in more detail, namely his argument from gunk. I will examine to which extent Schaffer's argument is convincing and conclude by discussing possible objections. Based on an intensive study of Schaffer's work, this paper aims to provide objections against Schaffer's argument from gunk and to grant a deeper insight into sparse ontology.

2. Priority Monism and Atomless Gunk

Priority monism is the view that there may be many concrete entities, but only one, the cosmos, is ontologically prior to all of them. Here, ontological priority is an asymmetrical dependence relation that holds between entities in general and not just facts in particular. Thus, monists regard the cosmos, a unified whole, as the only fundamental entity, a concrete entity that does not depend on any other concrete entity. In contrast, a pluralist might take the parts prior to the whole while characteristically arguing that the smallest parts are absolutely prior. Both parties may agree on the same number of existing entities but differ on the question of what is fundamental.

One of the key arguments in favor of monism is that it allows for the possibility of gunky objects, whereas pluralists cannot. In mereology, a gunky object is a composite material object in which each part is said to have proper parts and finally lacks any ultimate parts; it is infinitely divisible. Because parthood is transitive, every part of a gunky object is gunk itself. A world is considered gunky if every entity in that world has proper parts and no ultimate parts. However, a world may not be completely gunky; it may have certain gunky regions or objects. One could envision the existence of both possible atomistic worlds and gunky worlds, as well as worlds in between, with both atoms and gunk (Sider 1993: 219).

According to Schaffer, there are several reasons why both monists and pluralists should account for the possibility of gunk in their theory: First, the possibility of gunk is logically consistent, at least concerning classic mereology (Schaffer 2010: 61). Given that these classical models of mereology give rise to certain metaphysical possibilities, it turns out that gunky objects are

metaphysically possible (Schaffer 2010: 61). Furthermore, gunk is conceivable, meaning we can envision a gunky world without internal contradictions (Williams 2006: 494). Moreover, gunk seems to be scientifically profound. To support this claim, Schaffer (2010: 62) refers to the nobel-prize winning physicist Hans Dehmelt, who theorizes about an endless regress of the sub-electron structure (Dehmet 1989: 8618-8619), and theoretical physicist Howard Georgi, who proposes that quantum field theories could form an infinite pile descending to "arbitrary short distances in a kind of infinite regression [...]" (Georgi 1989: 505). According to these theories of physics, there is further evidence for the claim that atomless gunk is metaphysically possible. Thus, both monists and pluralists who want to put forward scientifically serious and empirically open hypotheses must allow for the possibility of gunk.

Here, Schaffer sees an advantage of priority monism over priority pluralism. If gunky worlds are possible, the pluralist, who argues that wholes depend on their parts, appears to be in a difficult position. For in a world with gunky entities, she must say that a gunky object depends on each of its parts, each of which depends on its parts without an end. Since these chains of dependence descend endlessly, the pluralist cannot appeal to atoms or other minimal fundamental entities to ground all others. However, this would appear to leave all gunky entities ungrounded and conflict with the foundationalist requirement that there must be a basis of being from which all things derive. If the pluralist does not argue against the possibility of gunk, she needs to devise a strategy that allows her to include gunky objects in her theory.

In contrast, the monist can quickly provide a theory that includes the possibility of gunk. Because the whole is prior to its parts, objects can be infinitely divisible without endangering priority monism as a doctrine. Since the whole cosmos is the one fundamental entity that grounds all others, gunky objects can continually divide themselves further without encountering a bottom. After all, they are ontologically dependent on the whole that forms the very basis of their existence. For the priority monist, thus, it does not matter whether the world is atomistic, has ultimate parts, or contains a mixture of gunk and atoms; priority monism allows all these theories to be true.

Taking this into account, Jonathan Schaffer's argument, which is put forward in favor of priority monism supported by principles of modality, can be set up in the following way:

Premise 1: Either the ultimate parts or the ultimate whole have to be fundamental in every possible world.

Premise 2: Gunky worlds exist without any ultimate parts (so the ultimate parts cannot be fundamental in gunky worlds).

Conclusion: Therefore, the ultimate whole must be fundamental in every world.

Note that a similar reconstruction of the argument can be found in Schaffer (2018: § 3.2.3 and 2010: 61-65). Jonathan Schaffer's argument from gunk is valid. Still, the soundness of his argument and, thus, the truth of Premises 1 and 2 is not secured. The extent to which the overall argument and its premises are compelling will be discussed critically. The following paragraph, therefore, focuses on the question of what possible objections can be raised against Schaffer's argument from gunk.

3. Objections

What objections can be put forward against Schaffer's argument from gunk? This section presents four objections to Schaffer's argument, which relate to different aspects of his reasoning. The objections vary in strength since some have more and others less far-reaching consequences for priority monism. Covered here are objections that shed light on consequences if one accepts the possibility of atomless gunk, as well as objections that challenge said possibility.

a. Atomless junk

A philosopher convinced of the infinite divisibility of gunky objects, or at a minimum of its possibility, may well have the same view regarding infinite composability. Much like everything could consist of gunky objects that decompose infinitely into tinier and tinier pieces, all could be mereological

junk that composes endlessly into greater and greater wholes. Just as the possibility of endless descent can be used against pluralism, the possibility of infinite ascent can be used against monism.

According to Bohn (2009: 28), a world is junky iff. each object in it is a proper part. If junky objects are metaphysically possible, it follows that priority monism is not necessarily true. This is the case because junky objects endlessly compose into greater and greater wholes so that there cannot be a fundamental whole in a world containing junk. A priority monist is now in explanatory trouble: how can she integrate the possibility of junk into her account? She seems to face the same problems as the pluralist when the latter acknowledges the possibility of atomless gunk.

But can the possibility of junk be accepted for similar reasons as the possibility of gunk? Schaffer (2018: § 3.2.3) states that the possibility of junky objects or junky worlds should be ruled out in principle. However, one can question whether he is right in asserting so. Some authors (Morganti 2009: 280-283, or Mormann 2014: 86) also argued that junk is part of mathematical models, such as set theory. Additionally, junk is conceivable. This implies that we can envision a possible world where junk objects exist without internal contradictions. In fact, there may not be any significant difference between conceptualizing a junky world and a gunky world.

Given that, the possibility that there is no ground of being for fundamental entities at the bottom from which others derive is on the same footing as the possibility of endless upward complexity. Hence, neither pluralism nor monism would have an advantage over the opposing side, and Schaffer's argument from gunk would have little persuasive power. It would appear that neither the ultimate whole nor the ultimate parts must be fundamental in all possible worlds, so both positions would have to be ruled out or declared contingent.

b. Contingency of priority monism

By stating that either the ultimate parts have to be fundamental in every world or the ultimate whole has to be fundamental in every world, Schaffer consequently argues that either priority monism or priority pluralism

necessarily must be true. But why must one of these two theses be true in all possible worlds with necessity? We could instead understand priority monism as a contingent rather than a necessary thesis.

According to Tho (2019: 289), natural laws supervene on the facts that constitute a given possible world. The laws of physics are likely to vary between possible worlds because the facts are also going to vary between these worlds (Tho 2019: 289). The contingency of the laws of nature is inferred from the ontological nature of the world about which these laws are concerned (Tho 2019: 289). What could that mean if applied to gunk? It might well be that what grounds what fundamentally varies among metaphysically possible worlds just as physical laws do: there are worlds where atoms as fundamental entities ground all derivative entities, other worlds where gunky objects are grounded in a fundamental whole. Thus, in some worlds, priority pluralism turns out to be true; in others (especially in those worlds where gunky objects exist), priority monism appears to be true. Whether one of these two theories is true in a possible world depends on the nature of that very world. Just as it may be that Newton's laws apply in some possible worlds but not in others, depending on the world's specific nature.

With this in mind, allowing for gunky objects might mean committing to the claim that there are gunky objects in some worlds in which priority pluralism might turn out to be false. In other worlds where there are no gunky objects, it is precisely pluralism that could be true. A problem would only arise if someone were to assume that there must necessarily be atomless gunk in every possible world, but this does not seem like something Schaffer needs to imply, and one would have to justify why this should be the case.

c. Middleism

This leads us to another reason to reject Schaffer's argument from gunk. One way to refuse his argument is to assume that other grounds exist aside from the cosmos or the ultimate parts.

Controversies about what is basic often involve presumptions that what is basic has to be either at a high or a low level. Sara Bernstein (2021: 1065) alternatively argues for middleism, claiming that entities at a middle level are basic. Middleism regards the entities that exist beneath and above the middle level as derivative; they all depend on those at the middle level (Bernstein

2021: 1065). A fundamental middle level thus grounds all the derivative entities that lie above and beneath it (Bernstein 2021: 1066). Following Bernstein (2021: 1070), this kind of world is metaphysically possible as both the notion of fundamentality and the definition of grounding do not require that there has to be a top or bottom level from which all other entities derive. A middle level is simply considered the most fundamental one because it builds the other levels. Whether or not our actual world is built in accordance with Sara Bernstein's account of middleism, the consistency and potential of middleism should raise doubts about Premise 1 of Schaffer's argument from gunk.

Another reason one might want to consider in support of middleism is the metaphysical possibility of hunky worlds or hunky objects. According to Bohn (2018: 168), there are hunky worlds where neither the ultimate parts are fundamental nor the ultimate whole is fundamental. A hunky world is such that everything in it both has and is a proper part. A hunky world is, like a gunky world, incompatible with priority pluralism and, like a junky world, incompatible with priority monism. Priority monism might account for the possibility of gunk and gunky worlds. Still, it cannot accommodate the possibility of hunk and hunky worlds, and this also applies to the pluralist who can accommodate the possibility of junk and junky worlds but not of hunk and hunky worlds. A middleist, instead, has an advantage over these two positions. If one accepts the possibility of hunky objects and hunky worlds, she can accommodate them. And this applies to gunky objects, gunky worlds, and junky objects or junky worlds, too.

d. Restricted composition

We frequently talk about an entity being composed of some other entities. Thereby, a gunky object is an object composed of endlessly smaller parts that themselves are also gunk. But what conditions have to be obtained for objects to compose something? In the field of metaphysics, this question has been called *the special composition question* (van Inwagen 1990: 21) and addresses whether and under what circumstances a set of entities composes a single entity.

One promising way of arguing against Schaffer is rejecting the possibility of gunk by attacking the principles of composition, which seem essential to

Schaffer's argument. This can be done in several different ways. For instance, in *Material Beings* (1990), Peter van Inwagen explicitly takes the absence of pieces of gunk as a starting point for his mereological metaphysics. He claims that every material object is either an atom or an organism, whereas he considers mereological atoms as objects lacking proper parts (van Inwagen 1990: 142). In this sense, material objects never compose anything. They are mereologically simple and thus partless. In this sense, van Inwagen discusses how mereological atoms can constitute a living organism but rejects the existence of non-living complex objects. If this is the case, there are no such things as desks, seats, houses, continents, or worlds. The only classes of composites van Inwagen accepts are living things that compose something just in case their activities constitute a life.

As every object is a living organism or a mereological atom, it follows that there are no such entities as gunky objects or even gunky worlds. At the same time, van Inwagen's approach has even greater implications for priority monism. Because material objects never compose other entities, there is no such entity as a cosmos or a fundamental whole of which all other derivative things are part. Agreeing with van Inwagen means not only raising an objection to Schaffer's argument from gunk but challenging priority monism as a theory in its entirety.

4. Conclusion

The central thesis of priority monism posits that there is one fundamental entity that is ontologically prior to all others. Schaffer's argument from gunk suggests that priority monism has an advantage over priority pluralism when it comes to accommodating gunk. He argues that both monists and pluralists should consider the possibility of gunk in their theories, as it is metaphysically possible and scientifically meaningful. By exploring the concept of monism and gunk, this paper presented Schaffer's argument from gunk, outlining the premises and conclusion of the argument and providing potential objections to it.

Some concerns have been raised here. First, one who believes in the infinite divisibility of gunky objects might as well entertain the idea of infinite composability. If junky objects are metaphysically possible, a priority monist

must explain how they fit into her framework, akin to the challenges pluralists face when considering the possibility of atomless gunk. Moreover, it could be argued that priority monism might as well be a contingent rather than a necessary thesis. Drawing on the idea that natural laws vary across possible worlds based on differing facts, one could argue that grounding principles, including whether priority monism or pluralism holds, also vary.

Additionally, middleism challenges the assumption that fundamentality must be achieved at either the highest or lowest level of reality and suggests that alternative grounding structures could exist where the entities at a middle level are basic. Some philosophers might even reject the possibility of gunk by attacking the underlying principles of composition. For instance, Peter van Inwagen argues that there are either mereological atoms or living organisms, excluding complex objects such as gunky objects or worlds.

By reconstructing Jonathan Schaffer's argument from gunk and providing objections to it, this paper provides a deeper insight into sparse ontology and questions concerning grounding and fundamentality. The objections brought forward might also raise further questions about the nature of fundamentality and composition.

References

- Bernstein, Sara (2021) "Could a middle level be the most fundamental?", *Philosophical Studies* 178/4, 1065-1078.
- Bohn, Einar Duenger (2018) "Indefinitely Descending Ground", in: Ricki Bliss and Graham Priest (eds.) *Reality and its Structure*, Oxford: Oxford University Press, 167-181.
- Bohn, Einar Duenger (2009) "An argument against the necessity of unrestricted composition", *Analysis* 69/1, 27-31.
- Dehlmert, Hans (1989) "Triton, ...Electron, ...Cosmon ...: An Infinite Regression?", *Proceedings of the National Academy of Sciences* 86, 8616-8619.
- Georgi, Howard (1989) "Effective Quantum Field Theories", in: Paul Davies (ed.) *The New Physics*, Cambridge: Cambridge University Press, 446-457.
- Giberman, Daniel (2021) "Deep gunk and deep junk", *Synthese* 199/3-4, 5645-5667.
- Morganti, Matteo (2009) "Ontological priority, fundamentality and monism", *Dialectica* 63/3, 271-288.

Mormann, Thomas (2014) "Set Theory, Topology and the Possibility of Junky Worlds", *Notre Dame Journal of Formal Logic* 55/1, 79-90.

Schaffer, Jonathan (2018) "Monism", in: Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2018 edition, <https://plato.stanford.edu/archives/win2018/entries/monism/>.

Schaffer, Jonathan (2010) "Monism: The Priority of the Whole", *Philosophical Review* 119/1, 31-76.

Schaffer, Jonathan (2009) "On What Grounds What", in: David Manley, David J. Chalmers and Ryan Wassermann (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press, 347-383.

Sider, Theodore (1993) "Van Inwagen and the Possibility of Gunk", *Analysis* 53/4, 285-289.

Tho, Tzuchien (2019) "The Immanent Contingency of Physical Laws in Leibniz's Dynamics", in: Rodolfo Garau and Pietro Omodeo (eds.) *Contingency and Natural Order in Early Modern Science*, 289-316.

Trogon, Kelly (2017) "Priority monism", *Philosophy Compass* 12/11, 1-10.

van Inwagen, Peter (1990) *Material Beings*, Ithaca, New York: Cornell University Press.

Williams, J. Robert G. (2006) "Illusions of gunk", *Philosophical Perspectives* 20/1, 493-513.

From Observable Regularities to Rule-Governed Practices: Exploring the Normative Facet of Social Reality

Amadeusz Just (Warsaw, Poland)

Abstract

The idea of the rule-governedness of social reality has been articulated for the first time by Peter Winch in his 1958 work, *The Idea of a Social Science and Its Relation to Philosophy*. Drawing on Wittgenstein's rule-following remarks and Rhees's critique of Wittgenstein's vision of language, he argued that the semantic and normative significance of human behavior is rule-dependent. The concept that social life is constituted by rules remains prevalent today, particularly among practice theorists. However, these theorists focus on the normativity of practices. They claim that regular sequences of actions are on their own bereft of any normative significance. What needs to be added – on their view – to render a regular sequence of actions normative are rules. Either a regular sequence of actions must be hinged on a rule to have normative significance, or it is only an observable regularity — that is, an empirical, nonintentional description of behavior that essentially lacks any normative dimension. In my talk, I shall challenge this conception by demonstrating that it relies on an implausible idea of internal or immanent rules. Insofar as any regular sequence of actions can be agreed with some rule, proponents of this view lack a criterion to distinguish between regular sequences of actions and rule-governed practices beyond explicit rules. But (1) not all regular sequences of actions that we would be willing to recognize as practices are bound up with explicit rules, and (2) not all human and non-human behavior that we would be willing to recognize as normative is guided by explicit rules. Thus, the rule-governed model of action may be helpful in describing only a restricted area of the normative facet of social reality, and rule-dependence is not the key to the normative significance of social life.

1.

In “Wittgenstein on Rules and Platonism” David Finkelstein discusses “a kind of Platonism about meaning and understanding” in Wittgenstein and some interpretations of Wittgenstein's rule-following remarks (Finkelstein 2000: 53). He gives an example of the kind of Platonism he has in mind by referring to §454 of PI: “How does it come about that this arrow --> *points*? Doesn't it seem to carry within it something extraneous to itself? – ‘No, not the dead line on paper; only a mental thing, the meaning can do that.’ – That is both true and false. The arrow points only in the application that a living creature makes of it. This pointing is *not* a hocus-pocus that can be performed only by the mind.” (PI 2009: §454). In this remark, the platonic stance is taken by Wittgenstein's interlocutor. He approaches the arrow as detached from the applications we make of it and sees it as an ink mark on paper. Abstracted from the character of life the arrow is only a dead sign bereft of any

significance. The interlocutor assumes that such a dead sign requires some kind of supplementation to have a meaning; something must be added to the inert sign of the arrow to make it *point*. Finkelstein comments that “[i]n order to explain how such a thing [‘dead line on paper’] could have a kind of significance that random squiggles lack, he [Wittgenstein’s interlocutor] imagines something *behind* the line on paper – something hidden – which, as it were, infuses it with significance” (Finkelstein 2000: 53). In §454, Wittgenstein’s interlocutor thinks that what needs to be added – or, in Finkelstein’s terminology, what is hidden behind the line on paper – is ‘a mental thing, the meaning’. That ‘mental thing, the meaning’ postulated by the interlocutor plays a twofold role in how he sees and wants to answer the philosophical difficulty. First, only by grasping the meaning behind the ink mark on paper we can *understand* that it is an arrow pointing in some direction. Otherwise, any attempt at giving an interpretation of the ink mark would result in nothing more but another bunch of dead signs or noises in need of yet another interpretation, *ad infinitum*. Second, only because of the mental entity behind the ink mark on paper the course of action – our looking or going into the direction pointed at by the arrow – is *determined*. Otherwise, nothing fixes the course of action, there is nothing to provide the standard of correctness. From the interlocutor’s perspective, without postulating the ‘mental thing, the meaning’ we are neither able to understand the rule, nor are the steps, that should be taken to follow the rule correctly, determined.

I am interested here in the interlocutor’s platonic stance pointed out by Finkelstein. Finkelstein argues that the interlocutor’s Platonism is “a desperate and unsatisfactory attempt to avoid the regress of interpretations which gives rise to the paradox of *Investigations*, §201” (Finkelstein 2000: 56). He also claims that the character of Wittgenstein’s criticism of Platonism is often misunderstood by the commentators as they “read Wittgenstein as endorsing one or another position that participates in the very thing to which he objects in platonism” (Finkelstein 2000: 54). I have already briefly discussed the former. The latter is exemplified by Saul Kripke’s reading of Wittgenstein and Crispin Wright’s answer to that reading. They both share with the Platonist his philosophical anxiety that there is a gap between a rule and its understanding (or meaning, or application) that must be bridged. However, instead of postulating some entity behind the statement of the rule, they give two

different, non-platonic answers to that anxiety. Finkelstein finds both of those answers unsatisfactory, but his main point is that from Kripke's and Wright's perspective it appears as if Wittgenstein were accepting his interlocutor's platonic anxiety as a real threat that should be addressed. Finkelstein argues that what Wittgenstein questions in the rule-following sections – especially in the second paragraph of §201 – is the very idea that “every rule comes with such a gap”. On his reading, Wittgenstein is trying to make us recognize that “when rules are seen as situated within our lives, such gulfs are exceptional. In general, nothing bridges a gulf between a rule and its application because no gulf opens up” (Finkelstein 2000: 64, 69).

Based on Finkelstein's discussion, two steps can be distinguished that are characteristic of the platonic stance. First, the Platonist approaches the rule as abstracted from the meaningful use it has in the context of life. Second, he assumes that such a rule – that now appears to him as mere marks or noises – requires some kind of supplementation. A fool-blooded Platonist takes the third step and postulates some kind of metaphysical entity that he thinks must be hidden behind the rule to fix its significance. At this point, a quasi-Platonist parts his ways with the full-blooded Platonist and – sharing with him the philosophical anxiety – looks for some more available grounding of the rule. In the next section, I will demonstrate that in her discussion of the normativity of practices, Jaeggi follows in the footsteps of Platonists.

2.

In her *Critique of Forms of Life*, Jaeggi championed an idea of immanent criticism of forms of life. She characterized forms of life as “ensembles of social practices” (Jaeggi 2018: 55). Immanent criticism is directed not at forms of life in general but at internal norms or rules embedded in social practices. Therefore, its possibility depends on Jaeggi's preferred social ontology according to which rule-governed social practices constitute the internal texture of forms of life. Jaeggi elucidated her understanding of the concept of practice by discerning its seven aspects. Even a brief presentation of all of them would require a more extended discussion. Therefore, I shall move directly to the fourth aspect according to which practices are rule-governed.

Jaeggi claims that the “formation and performance of practices involve more than just observable regularities. They always involve sequences of actions governed by rules and regulations, hence by a division of the possibilities of action into what is and what is not appropriate to do”. At this point, she refers to John Rawls’s definition of a social practice. I will come back to this. She continues by quoting Titus Stahl – who proposed his own idea of internal critique – saying that “[t]he central idea is that a practice invariably involves an internal distinction between right and wrong action.” Jaeggi comments that “[t]he decisive point is that the operative criteria are *internal to practice*. If one can act wrongly in different ways with respect to practices, then this is a matter of rule violations that miss the point of the practice itself” (Jaeggi 2018: 57).

This passage quite clearly shows that Jaeggi’s goal in postulating the rule-governedness of practices is to establish their normativity. For this purpose, she contrasted rules with observable regularities and claimed that observable regularities are not enough to constitute the normativity of practices. She didn’t give at this point any reason for that claim, nor did she explain her understanding of the term observable regularities. Given that, it will be helpful to turn in this respect to Stahl. Stahl agrees with Jaeggi that regularities of behavior are not sufficient in characterizing social practices and their normativity. What distinguishes practices from such regularities are – according to him – “rules that determine the *correctness* and *incorrectness* of certain forms of behaviour” (Stahl 2022: 162). Fortunately, Stahl is more generous than Jaeggi and provides the reader with some background for that claim. He refers to Saul Kripke’s reading of Wittgenstein and says that “any descriptively graspable *regularity*” (Stahl 2022: 177) would not constitute a fact acceptable for the skeptic as determining the meaning of the rule. Insofar as this is true in reference to Kripke’s skeptic, who is interested in establishing the meaning of a *rule*, it is not clear what consequences it should have for Stahl or Jaeggi, who are interested in establishing the normativity of a *practice*. Anyhow, Stahl seems to agree with Kripke’s skeptic since – in some other parts of his book – he refers to observable regularities as empirical, nonintentional descriptions of behavior that essentially lack any normative dimension (cf. Stahl 2022: 162, 177ff, 185). They are empirical facts bereft on their own of any normative significance. It seems that Stahl assumes that regularities, as

something that can be plainly observed, are – so to say – empirically given. I cannot discuss it here, however, I want to emphasize that for Wittgenstein grasping a regularity is something that requires a technique (cf. e.g., RFM 2001: VI, §2) and already this has significant implications for the issue of normativity.

If my reconstruction of Jaeggi's account of observable regularities in reference to Stahl is correct, then Jaeggi approaches social practices as mere sequences of actions that, even though regular, require some kind of supplementation to gain normative significance. Given that, any description of observable regularities in the people's movements and actions would be a description devoid of any reference to normativity. In the previous section, I have discussed Wittgenstein's interlocutor platonic stance. Finkelstein characterized a Platonist about meaning and understanding as "someone who, in an effort to explain how mere noises and marks can have semantic significance, is driven to posit self-standing sources of significance – items which stand to the significance of our dead marks and noises as the sun stands to the light of the moon" (Finkelstein 2000: 53-54). My argument is that Jaeggi, in an effort to explain how mere regular moves and actions can have normative significance, is also driven to posit self-standing sources of significance – rules. Someone could argue that while in the case of Platonism about meaning and understanding Wittgenstein's interlocutor postulated some mysterious entities behind the mere signs and noises without giving any (convincing) account of what they might be, in the case of practices, the entities postulated behind the sequences of actions are not mysterious at all. They are rules. Indeed, there is a non-mysterious account of rules constituting the normativity of practice in John Rawls's paper "Two Concepts of Rules" to which Jaeggi referred.

Rawls defined a practice technically as "any form of activity specified by a system of rules which define offices, roles, moves, penalties, defenses, and so on, and which gives the activity its structure" (Rawls 1955: 3). According to Rawls, rules constitute both the meaning and normativity of a practice. However, there is a significant difference between Rawls and Jaeggi in understanding what a *rule* is. For Rawls, "[i]t is the mark of a practice that being taught how to engage in it involves being instructed in the rules which define it, and that appeal is made to those rules to correct the behavior of those who engage in it. Those engaged in a practice recognize the rules as

defining it. [...] it is essential to the notion of a practice that the rules are publicly known and understood as definitive” (Rawls 1995: 24). The possibility of being instructed in the rules or corrected in one’s behavior by an appeal to rules requires that the rules of practice are publicly known and, therefore, explicit. In contrast, Jaeggi’s primary interest lies in what she calls ‘internal’ or ‘immanent’ rules. There is nothing philosophically troublesome in saying that the normativity of the chess game is constituted by the rules of the chess. The normativity of the game of chess does not require postulating any internal rules for that practice. There are explicit, publicly available rules of chess. (Wittgenstein in the rule-following remarks went further and asked a question concerning the constitution of the normativity of rules.) But Jaeggi is not interested in such obvious cases. She claims that all practices are rule-governed and it seems that – like Rawls – she models her understanding of rule-governedness of practices on cases like the game of chess. However, there is a significant difference in the role rules play in practices between practices like a chess game and examples provided by Jaeggi.

One of Jaeggi’s examples of social practice is shopping. She writes that “*depending on how one understands* the practice of shopping, talking loudly on one’s cell phone while in the store or leaving the store without saying goodbye can constitute an internal violation of the norms associated with shopping” (Jaeggi 2018: 58, emphasis added). This is a striking claim if we remember – that according to Jaeggi – what constitutes the normativity of practices are rules. It is not the case in reference to the chess game that ‘depending on how one understands’ that practice, such-and-such a move *can* constitute a violation of the game. What can depend on someone’s understanding of chess is e.g. a strategy she will choose. A person with a better understanding of such a game usually is a better player. But what counts as a correct or incorrect move in chess does not depend on one’s understanding of that game. It depends on the rules of chess. Similarly, traffic violation does not depend on our understanding of the practice of driving a car but on the traffic rules. If someone runs over a red light, we would not say that he understands the practice of driving a car differently. He either doesn’t know the traffic rules or breaks them intentionally. I agree with Jaeggi that ‘leaving the store without saying goodbye’ can be considered rude depending on how one understands the practice of shopping. However, one can think of many

scenarios why the person didn't say goodbye and, in some of those scenarios, not saying goodbye would not be considered rude. In the practice of shopping, even such a simple thing as saying or not saying goodbye has a complex normative significance depending on the context of a particular case. While in the case of rules – like the rules of chess – their normativity is not in that way context-dependent. Insofar as leaving the store without saying goodbye could be satisfactorily excused or explained, there is no explanation for diagonally moving your rook in chess. Even if you would excuse such a move by saying that you mistakenly took your rook for a bishop, such a move would still be incorrect. And that incorrectness is constituted by the rules of chess.

Let me get back to the contrast between observable regularities and rules. I have suggested that Jaeggi approaches regular sequences of actions as bereft of normative significance. Borrowing from Wittgenstein, it can be said that it appears to her that there is a gap between a regular sequence of action and its normative determination. She thinks that what is needed to bridge that gap is a rule. Later in her book Jaeggi gives an example that illustrates my argument. She imagines herself as someone who always walks without stepping on the lines between the paving stones and says that this exhibits merely regular behavior. If she stepped one day on the line between paving stones, “then I – Jaeggi comments – have done something *different* from usual, but I have not done anything *wrong*.” She continues that only if she would make it a rule not to step on the lines, stepping on the lines would be “a violation of this rule”. In the end she comments that “[o]nly when a certain regular behavior is required, so that one can also violate this requirement, are we dealing with norm-governed behavior” (Jaeggi 2018: 93).

The stake of the contrast between observable regularities and rules was to demonstrate that the former are on their own non-normative. Jaeggi described two cases of someone who usually walks without stepping on the lines between the paving stones, but steps on a line one day. The only difference between those two cases was that in the second one the person formulated a rule – not to step on the lines – which she was following. It is clear that only if there is a rule we can say that someone violated that rule. But it does not demonstrate that the only possible kind of requirement for our behavior has a form of a rule. Meanwhile, it is precisely this train of thought that pressures Jaeggi to postulate rule-governedness of practices. She thinks that only rules

can formulate requirements for our behavior and constitute normativity of practices. Therefore, if a sequence of actions is supposed to be taken as social practice, it has to be supplemented with rules. If there are no explicit rules that can be linguistically stated, then there *must* be hidden some internal rules.

But the trick in Jaeggi's argument lies in its first-person scenario: I can only do something wrong if I am following rules in what I'm doing. However, if we switch to a third person, how can we distinguish between regular behavior of someone who usually walks without stepping on the lines between the paving stones and rule-governed behavior of someone who usually walks without stepping on the lines between the paving stones? The only available criterion is the explicit rule 'do not step on the lines between the paving stones'.

3.

In this paper, I have tried to reconstruct Jaeggi's reasoning that led her to the postulate of the rule-governedness of practices. In her account she approaches practices as bereft of normative significance. It appears to her that there *must* be something behind mere sequences of actions to constitute a normative practice. She thinks that what lies behind observable regularities of actions are rules. Insofar as we think about rules as explicit and publicly available, there is nothing philosophically problematic in referring to rules as constituting the normativity of practices. However, Jaeggi's immanent criticism is directed not at explicit, but at *internal* rules embedded in practices. I tried to show that Jaeggi's postulate of internal rules expresses a misunderstanding of the role that rules play in our social practices. She is led to a confusion regarding what actually is called rules by reasoning very similar to the reasoning we could see expressed in the platonic stance of Wittgenstein's interlocutor.

References

Finkelstein, David H. (2000) "Wittgenstein on rules and platonism", in: Alice Crary and Rupert Read (eds.) *The New Wittgenstein*, London and New York: Routledge, 53-73.

Jaeggi, Rahel (2018) *Critique of Forms of Life*, transl. by C. Cronin, Cambridge: The Belknap Press of Harvard University Press.

Rawls, John (1955) "Two Concepts of Rules", *The Philosophical Review* 64/1, 3-32.

Stahl, Titus (2022) *Immanent Critique*, transl. by J.-B. Oduor, Lanham: Rowman & Littlefield.

Wittgenstein, Ludwig (2009) *Philosophical Investigations / Philosophische Untersuchungen*. P.M.S. Hacker and J. Schulte (eds.), transl. by G.E.M. Anscombe, P.M.S. Hacker and J. Schulte, New York: Wiley.

Wittgenstein, Ludwig (2001) *Remarks on the Foundations of Mathematics*, G. H. von Wright, R. Rhees, G.E.M. Anscombe (eds.), transl. by G.E.M. Anscombe, Oxford: Basil Blackwell.

Acknowledgement

This paper is a result of the research project “Ludwig Wittgenstein and Social Normativity” (no. 2019/33/N/HS1/03103) financed by the National Science Center, Poland.

Quasi-objects

Christian Kanzian (Innsbruck, Austria)

Abstract

My talk contributes to the first section “Reality and Cognate Notions” by considering the concept of “quasi-object” as a citizen of the grey zone of “cognate notions”. I will argue that the theoretical objects or posits of natural science, including quantum physics, are such quasi-objects. This contrasts with the view that they are entities, which constitute our everyday life-world from the “bottom up”.

I begin by introducing the concept “quasi-object”, bringing together some of Eli Hirsch’s metaontological considerations with Jonathan Lowe’s ontological interpretation of the term (section 1). Then, in a brief excursion into the philosophy of science, I ask what sciences do when they introduce quasi-objects into their theorizing (section 2). I then return to the problem of what really exists – if not quasi-objects (section 3) – and give a brief outlook on some consequences for realism in philosophy of science and philosophical naturalism.

The main focus of my considerations is ontology and metaontology, and specifically the “objects” of natural science. The other theorems play the role of supplements and thus have a provisional and hypothetical character. In these respects, of course, my results, too, remain provisional and hypothetical.

1. What quasi-objects are

As mentioned in the summary, my introduction to the concept “quasi-object” begins with some metaontological considerations. They concern the distinction between an ontologically *proper* and ontologically *improper* way of speaking, especially in “there is/are...”-claims, as we find it in Eli Hirsch’s theory of “as if”-quantification (cf. Hirsch 2002). This is sometimes represented in the literature by differentiating between external and internal quantifiers (cf. Hofweber 2005) or heavyweight and lightweight quantifiers (cf. Chalmers 2009).

The common conviction behind this distinction, however, is that not every use of variables x , y , ..., combined with a “there is/are ...”-claim, automatically commits the speaker to accept the existence of the referents of the expression. Hirsch’s basic idea is that we can also speak about something *as if* it were an existing entity. This does not *solve* ontological puzzles, but it can help to “deflate” the problems behind them. A paradigmatic problem that Hirsch has in mind is the endurer/perdurer debate, which has heavily shaped the contemporary ontological scene. Are there three-dimensional diachronically

identical things or substances, bearers of properties and changes – endurers; or are there only four-dimensional sums of numerically distinct spatio-temporal parts or phases – perdurers? According to Hirsch, as endurantists, it is not necessary to refute perdurantism as such, since we can legitimately allow there is/are-claims to run over four-dimensional objects or perdurers, as an improper or as-if-use of the quantifier \exists . As endurantists, we can speak *as if* there were perdurers, if we are prepared to give an account of the theoretical context of our liberalism.

Important for us is that the findings of quantum physics may provide such a theoretical context, since four-dimensionalism is more successful for physical modellings than any three-dimensional thing- or substance ontological frame. Quantum-physical models demand a language in which we speak about four-dimensional objects. Why should we negate that and exclude a grammar in which we speak as if there really were such perdurers?

Another as-if way of speaking in line with the results of quantum physics makes use of particular properties or fields, without determinate identity, called *tropes* (cf. Campbell 1990, chapter 6). Some who are open to the notion of tropes speak of ordinary objects, traditionally categorized as things or substances, as if they were bundles of tropes, interdependent in their existence. Hirsch would invite substance ontologists to be tolerant and to accept trope talk as legitimate as-if talk when it is helpful in some theoretical contexts.

According to Hirsch, it is not only possible but also advisable to integrate such improper ontological talk into a proper or ontologically serious way of speaking. As tolerant endurantists, we should integrate perdurer and trope talk into the grammar of an endurer language. Then we could speak – because of some theoretical constraints – as if things or substances were perdurers or even trope-bundles. We could give them names, predicate attributes of them, and even count them if we allow the introduction of alternative principles of individuation.

Here we arrive at precisely what Jonathan Lowe and others have called *quasi-objects* (cf. Lowe 1998, 58, 70ff). Quasi-objects are referents of theoretical concepts whose identity remains undetermined for principal, not contingent,

reasons. We cannot say what the identity of quasi-objects consists in. We cannot provide identity criteria for them. Nevertheless, we can count them, which presupposes that we are able to individuate them in a reasonable way.

According to Lowe, the theoretical objects of quantum physics are the paradigmatic examples of such quasi-objects. The posits of the standard theories of quantum physics have no determinate identity, which does not prevent scientists from defining metrical orders that allow these objects to be counted. Lowe calls them “esoteric objects” in a non-pejorative sense (ibid, 58). Their acceptance is legitimate only *within* a particular theoretical context.

To allow something without determinate identity would put us ontologists in an awkward or at least difficult position if we could not find support in Hirsch’s as-if-quantification: even from the point of view that there are – in the proper sense – only entities with a determinate identity, we can allow that, for some theoretical reasons, e.g. in quantum physics, we can speak *as if* there were the aforementioned “esoteric objects”. We just have to allow ourselves to use a non-ontologically committing quantification, one that is lightweight or, for friends of Carnap, “internal”.

In short: Quasi-objects are the referents of non-ontologically committing there is/are-claims; they belong to the scope of lightweight/internal quantifiers that we can use *as if* we were speaking about real objects. Quantum-physical quasi-objects are “esoteric” in the sense that they can only be legitimately mentioned *within* or *internal to* a given theoretical context, even allowing for identifiable objects without identity.

To classify the objects of quantum physics as (perhaps not the only but a representative genus of) quasi-objects, as Lowe has suggested, presupposes not only a remarkable set of philosophical assumptions but also and chiefly a specific understanding of scientific practice. In the next section, I want to give a brief survey of this view of scientific practice.

2. What sciences do

The most relevant aspect of the presupposed understanding of scientific practice is that it is a methodological interpretation of empirical data; and that

for this interpretation it is significant to apply *models*. Crucial for our context is that such interpretations, and the choice of models, do not allow any decision to be made about what there is, or what exists. Scientific theories bring empirical data into a systematic order, but they do not explain reality. (I think this is also the core idea of 6.341 in Wittgenstein's *Tractatus*.)

How can we make this a little more plausible? Without claiming to have any detailed knowledge of atomic physics, it is clear that scientists do not come to their theories by simply picturing some (putative) objects. Their talk of atoms or subatomic particles is the result of an interpretation of empirical data, in which models obviously play an important role. Take for example the planetary model of atoms, according to which some electrons orbit around a nucleus composed of neutral and positively charged particles. This model has important theoretical advantages. As we all learnt at school, it helps us to get along with some specific experimental data, such as Rutherford's scattering. But the acceptance of the model, and this is the crucial point, does not commit us to the strange assumption that at a certain level of the micro-world, there are actually some entities moving around, just like the planets of our solar system. The same can be said in respect of models in quantum physics, as some theories mention waves or a wave/particle dualism. These are theoretically valuable models that interpret empirical data.

I follow those experts in the philosophy of science who do not see the aim of scientific theories as providing a description or elaboration of a picture of reality (see e.g. Drieschner 2021, 78 "Was man von der Quantentheorie bekommt, ist nicht eine Beschreibung der Wirklichkeit, [...]"; my translation: What one gets from quantum theory is not a description of reality). Science does not exhibit entities, nor does it allow any implication about what there is or what exists, in an ontological sense. Scientists try to provide a system of propositions from which we can infer reliable prognoses of possible measurements (cf. *ibid*). The main means of such a methodological systematization of propositions is the application of models.

We can also find support from authors who have dealt with the concept of models in relevant scientific contexts, including, among others, Tarja Knuuttila and Marvin Rost. They regard models as "epistemic artefacts" (see: Rost and Knuuttila 2022), tools that aim to provide theoretical explanations, and

conditions for applying theories to empirical data, or to verify/falsify them empirically. Rost and Knuuttila also explicitly warn against simplifying picturing theories of scientific models. Models do not describe reality. They do not provide insight into the basic structure of reality. Scientists do not picture the world.

Although this is dangerous for one of the philosophical goals of this paper, anti-naturalism, we can also refer to Quine and his concept of *posits* of scientific theories. Quine's examples of posits in *Word and Object* are molecules, chemical compounds (Quine 1960, § 6, 21f). Quine does not call them “quasi-objects” but rather “extraordinary things”, which are introduced relative to the requirements of a theory to bring systematic order to a complex diversity of empirical data. Posits or extraordinary things are only relevant *within* the context of the theory. It is also clear in Quine that a naive realistic picture theory, e.g. about molecules, must fail. We introduce posits, e.g. molecules, because it helps us to interpret empirical data according to some models in physical, or better *chemical*, theories. Elsewhere Quine calls his posits “convenient intermediaries” between conceptual schemes and the empirical data we obtain in scientific experiments, in order to bring them into a practicable form that allows prognoses (cf. Quine 1951, 41). This fits neatly with his remarks about extraordinary things in *Word and Object*. Posits are not real objects; rather, they correspond to models that allow a methodological interpretation of empirical data, in the most economical, simple, and consistent way. Posits are quasi-objects, in the sense introduced.

It is particularly relevant for our context that even in the philosophy of science, authors concede that, in order to achieve the explanatory goal, it may make sense to speak of, e.g. posits of models in quantum physics *as if* we had to do with real objects, to which we can attribute some properties, and which can undergo change. We can talk about “extraordinary things” (Quine) as quasi-objects, as long as we do not forget that they are only “esoteric”, in Lowe's sense.

This is also the point of our excursion into the philosophy of science. If the understanding of scientific practice sketched here is correct, then the putative objects of, e.g., quantum physics are in fact posits of models that can be treated quasi as real objects. This is, to put it cautiously, compatible with some

moderate versions of scientific anti-realism, which can, of course, be criticized and questioned. But, as far as I can see, it stands in a line of good traditions, not only in analytic but also in continental philosophy. As long as the alternative, scientific realism, remains far away from final victory, we as ontologists dealing with quasi-objects can sleep well with the hypotheses about scientific practice referred to above.

3. What really exists

In order to integrate quasi-objects into an ontological scheme, we must be ready to systematically connect them with some really existent beings, not-only-quasi-beings. That raises the following question: What is it that really exists? In the language of metaontology: What are really committing existential quantifiers, external, heavyweight?

Eli Hirsch has a suggestion that I would like to take up here. He actually assumes a basic, non-as-if-meaning of the existential quantifier, in consequence of existence, a linguistic vehicle that really commits ontologically.

For those less familiar with descriptive metaphysics or ontology, beginning with Aristotle and represented by Peter Strawson (1959, introduction) in contemporary debates, this may sound a little surprising. Not so to others: It is not the idealized language of natural science but our normal conceptual scheme that covers the basic structures of reality. So, the real committing quantifier is the existence quantifier of plain English. It is our ordinary use of there is/are-propositions that commit ontologically.

But what we primarily commit ourselves to with this “master quantifier” in plain English are macroscopic, three-dimensional, diachronically identical objects, enduring things or substances. The priority of commonsensical (linguistic) practice initially leads to the acceptance of an endurer-ontology. From the perspective of metaontology, we can say that thing- or substance-existence is the proper mode of existence.

This provokes an objection: is this not an oversimplifying deduction of categorical schemes from ordinary linguistic practice? I must confess that I am not able to fully reject this objection here. Just let me express my conviction:

No, not at all, if we are ready to develop a systematic theory of such endurers, and “endurer quantification”, avoiding all kinds of oversimplification, in (meta)ontology as in semantics. (For an attempt, see Kanzian 2020)

What I can do here is to present some additional arguments in favour of the aforementioned priority thesis, derived from what I have said so far. First, on the level of metaontology: Here we can say that in fact we need a basic meaning of existence and of the existential quantifier. If we were to regard all meanings as equal, we would end up in a “weird form” (Hirsch) of quantifier variance without limits to linguistic idealism.

Then, from an ontological point of view, a first step towards an endurer-based categorical scheme: What exists must have *identity* and *individuality*, as we learn from Lowe, and others, of course: No entity without identity. Endurers, whatever they are, are in any case identical with themselves, not only diachronically but also across possible worlds. If we cannot provide reliable conditions of identity for something, it cannot be an endurer. And endurers are countable; they have stable principles of individuation. Of course, there are also other candidates of identical *and* identifiable entities, such as the properties characterizing things or substances, and the changes and events in which they are involved. These, in contrast to quasi-objects, also exist but, as can be shown, in an analogous or dependent way. Endurers are not the only existents, but the *prote ousia*, the primary beings.

Furthermore, a brief semantic remark in favour of the priority of the endurer language: According to Hirsch, it is the endurer language, the language with primary endurer quantification, which allows an integration of alternative linguistic forms, including propositions with there is/are-phrases running over four-dimensional perdurers. It is even possible to define truth-conditions for these semantic alternatives in the endurer language; in a systematic way, by referring to the possibility of as-if-talk about quasi-objects, as I have suggested in the previous sections. The reverse is not the case. There are no successful reconstructions of an endurer language from the basis of a perdurer language. I know this sounds like a bold claim, but I think it is defensible.

The last line of argumentation I want to mention briefly here comes from the philosophy of science, and it brings our quasi-objects in connection with the

real or non-quasi-beings, enduring things. In this context, I can offer two strategies. As to the first strategy: Real endurers are presupposed methodologically in the modelling of science, including quantum physics. For this I can refer to a reliable witness in the form of Albert Einstein, who explicitly concedes that without the idea of independently existing macro-things, which we have from our commonsensical, endurer-dominated macro-world, model-based thinking, even in quantum physics, would be impossible (cf. Einstein 1948).

The standard examples of models I mentioned earlier should also make this clear: atoms modelled as planets orbiting a star, and the talk of waves and particles in quantum theory. All these models are derived from our endurer-world. In this respect, the quasi-objects posited in these models depend on the non-quasi-objects of our commonsensical life-world, including real planets, waves, and, of course, particles. This can legitimize the assumption that our everyday life practice is fundamental relative to scientific practice; that the latter is a methodologically limited derivation of the former.

The second strategy is also a way of bringing together endurer-things and quasi-objects in a consistent ontological theory, avoiding the shortcoming that atoms, subatomic particles, and waves have no connection with reality at all. What is true is that endurers exist. But whatever endurers are, they have, as parts of our commonsensical life-world, a *material* component. This material component can be analysed in an ontological way, insofar as it plays an important role in the composition of the complex unity of things and substances. But matter can also be analysed from the perspective of various natural sciences. Then, as we have assumed, models are indispensable. We can talk about the posits of these models *as if* they were things or substances in themselves. This is the theoretical locus of our quasi-objects, which brings us back to the main topic of this talk and shows that they can be brought in systematic connection with the non-quasi-objects.

Quasi-objects are the posits of models that support the scientific interpretation of the material component of real entities.

4. Consequences

The consequences should be not too hard to see: About the implications for the philosophy of science I have said enough. The other implication is that of anti-naturalism. This is not especially surprising, but it can be made more explicit. If it is true that the (putative) particulars of the micro-world that quantum physics presents to us are not entities but quasi-objects, then it is false that quantum physics can tell us what there is. Natural science is not an authority for explaining reality, which is what naturalistic ontologies assume. In other words, ontology is not an “a posteriori” discipline dependent on quantum physics.

If we want to interpret our world ontologically, we should be open to starting with our commonsensical life-world, with its real existing inhabitants, things or substances, and all the other entities that depend on them: the properties they possess and the events in which they participate.

This is not, of course, an attack on natural science. Scientists do an excellent job. They interpret empirical data in a methodical way, successfully in both theory and practice. This is a plaidoyer against philosophical naturalism, insofar as it misinterprets science and treats its quasi-objects as entities; it confuses reality with mere cognate notions.

References

- Campbell, Keith (1990) *Abstract Particulars*, Oxford: Blackwell.
- Chalmers, David (2009) “Ontological Anti-Realism”, in: D. Chalmers et al. eds. *Metametaphysics*, Oxford: Clarendon Press, 77-129.
- Drieschner, Michael (2021) „Quantenmechanik verstehen“, *Information Philosophie*, September, 78-92.
- Einstein, Albert (1948) „Quantenmechanik und Wirklichkeit“, *Dialectica* 2, 320-324.
- Hirsch, Eli (2002) “Quantifier Variance and Realism”, *Philosophical Issues* 12, 51-73.
- Hofweber, Thomas (2005) “A Puzzle about Ontology”, *Nous* 39, 256-283.
- Kanzian, Christian (2020) *Alltagsontologie. Eine metaontologische Grundlegung*, Innsbruck: iup.
- Lowe, Jonathan (1998) *The Possibility of Metaphysics*, Oxford: Clarendon Press.
- Quine, Willard V.O. (1951) “Two Dogmas of Empiricism”, *Philosophical Review* 60, 20-43.

Quine, Willard V.O. (1960) *Word and Object*, Cambridge, MA: MIT Press.

Rost and Knuuttila (2022) "Models as Epistemic Artifacts for Scientific Reasoning in Science Education Research", in: *Education Sciences* 2022, 12, 276. <https://doi.org/10.3390/educi12040276> (15.7.2022)

Strawson, Peter (1959) *Individuals: An Essay in Descriptive Metaphysics*, London and New York: Routledge.

Acknowledgement

I thank Alois Pichler for his valuable comments on an earlier version of this paper.

Shared Responsibility Does Not Follow from Shared Agency

Pelin Kasar (Vienna, Austria)

Abstract

In this paper, I discuss *moral agency* and *moral responsibility* attribution in the context of Artificial Intelligence (AI) and machine learning advancements. The emergence of increasingly autonomous systems challenges the standard view of these two concepts. We are said to be faced with an ever-widening *responsibility gap*. One response to the responsibility gap is the *hybrid responsibility approach*, which suggests distributing moral responsibility across human and machine components. According to this view, joint responsibility follows from the theories of extended agency, where the agency goes beyond the human component. However, I argue against this approach because of the inference made from extended agency to extended responsibility. Instead, I advocate for Johnson and Power's surrogate agency model, which allows for *shared agency* between humans and machines without implying *shared responsibility*. Ultimately, I emphasize the importance of developing a model that accommodates joint action while preserving human moral responsibility in interactions with machines.

1. Introduction

The questions on responsibility attribution in the field of Artificial Intelligence (AI) have gained significant prominence in light of recent advancements, particularly in machine learning algorithms. The emergence of semi-autonomous, independent, self-learning systems, where human control is significantly diminished, poses a challenge to the traditional understanding of responsibility attribution. According to the standard view, human makers or users are held responsible for the consequences of using simple tools and complex machines. However, when humans lack sufficient control and knowledge of the consequences of machine actions, questions arise regarding how responsibility can be assigned. In response to this widening "responsibility gap" (Matthias, 2004), one compelling view is the concept of joint responsibility proposed by Allan Hanson. Hanson contends that since the agency is shared between human and machine components, responsibility for a machine's behaviour is distributed across a network of human and machine elements (Hanson, 2009). In this paper, I argue against the notion of hybrid responsibility as a plausible response to the responsibility gap. While acknowledging the possibility of certain machine-human actions being conceived as joint actions, I believe that moral responsibility ultimately resides with human beings, even in the context of autonomous machine actions.

2. The Standard View of Responsibility Attribution

The concepts of moral responsibility and moral agency have historically been theorized in relation to human beings. Humans are distinguished from other entities by their capacity for intentional action. Moral agency refers to the capacity to perform intentional actions, with moral responsibility primarily ascribed to the agent of such actions. While a falling rock may cause harm, it is not deemed a moral agent nor morally responsible, as it lacks the capacity for intentional action. According to the standard view, intentional action requires an agent with internal mental states, such as beliefs, desires, and intentions, as these mental states cause the agent's behaviour. Typically, in intentional actions, the agent has control over and awareness of their actions. Without these conditions fulfilled, moral responsibility is not typically attributed to the agent. Thus, to hold an individual (S) morally responsible for their behaviour (X), at least two conditions must be met:

1. S must be an agent possessing relevant internal mental states.
2. X must be an intentional action over which S has control and awareness.

Agents do not act in a vacuum; they deploy various means to achieve their ends. Often, human-made artefacts serve as necessary means to accomplish these ends, with responsibility primarily resting on the user as long as artefacts function as intended. However, not all artefacts are similar in terms of their capabilities. Autonomous machines employing machine learning, which inherently involves unpredictability, challenge traditional notions of responsibility attribution. Examples include self-driving cars, Eurotransplant, and ChatGPT, among others. These AI technologies can make decisions and act without direct human intervention, with their operational rules subject to change during use. The engineers who design and build such systems often have limited foresight into their eventual actions, leading to a lack of control over potential consequences. If designers and users seem to have no sufficient control over these machines, how are we going to hold them responsible? The independent and autonomous nature of certain technologies undermines the control condition and gives rise to what Matthias refers to as the responsibility gap (Matthias 2004).

David Gunkel (2020) surveys three main responses to the responsibility gap: instrumentalism 2.0, machine ethics, and hybrid responsibility. Instrumentalism 2.0 regards autonomous machines as mere instruments or tools, emphasizing that only humans possess moral responsibilities. While it is an appealing option with its focus on human responsibility, this approach fails to acknowledge the significant autonomy of certain AI systems, equating them with simple tools like hammers. We need to draw a distinction between them. In contrast, machine ethics draws that distinction very sharply; they consider the potential of the capacities of certain AI-powered machines to make ethical determinations. They go against the traditional understanding of responsibility attribution that requires agents with mental capacities. In the next section, I discuss an intermediary perspective between these two extremes: the hybrid responsibility approach.

3. Hybrid Responsibility Approach

The hybrid responsibility approach presents an intermediate response to the responsibility gap. According to this view, responsibility is distributed across a network of interacting human and machine components (Gunkel, 2020). Allan Hanson proposes that the concept of joint responsibility can resolve the problem of the responsibility gap. According to Hanson, joint responsibility follows from the theories of extended agency, according to which most actions are undertaken by subjects that extend beyond the human individual to include other human beings and any number or kind of nonhuman entities (Hanson, 2009). Central to his theory is the recognition that humans invariably rely on artefacts to achieve their goals. He proposes that if these elements are necessary for executing the action, they should be viewed as components of the agency for the action itself. For instance, when a blind individual uses a stick to navigate, Hanson contends that the composite of the individual, the stick, and the environment constitutes the agent (Hanson, 2009).

An important problem I see in Hanson's theory is that he seems to be mistaking the necessary conditions for the actions as constituents of the agency and, therefore, constituents of the moral responsibility. He takes the simple fact that humans need artefacts to achieve their ends as the basis of his extended agency theory. When Hanson says in the case of a blind man using a

stick to walk down the street, the agent is the composite of the man, the stick, and the street, we see that he takes the necessary conditions as the constituents of the agency. It is true that the blind man's action causally depends on the existence of the street and the stick, but is this enough to claim that they are part of the agency and, more importantly, the moral responsibility? It seems less problematic to attribute them causal responsibility and regard them as necessary enabling conditions for the action. When I drink a coffee with a coffee cup, the coffee cup enables me to drink the coffee, but in the end, I am the one who is doing the action. It causally contributes to the action, but this doesn't seem sufficient for taking it as part of the agency. Nevertheless, this could be different with increasingly autonomous machines. I will later talk about this possibility through the concepts of surrogate agency and joint action.

According to Hanson, from the extended agency, it is a short step to joint responsibility (Hanson 2009). He moves easily — almost too easily — from his theory of action to his theory of moral responsibility. For this step, he appeals, on his conception, the commonly accepted claim that moral responsibility for a deed lies with its doer. Since he thinks the doer involves human and non-human components, he argues that the responsibility is distributed across the human and non-human components as well. For instance, he claims that when Eurotransplant generates a priority list of recipients, the moral responsibility doesn't lie with human programmers and users alone; it could also lie with databases, computer hardware, and software. How should we understand the claim that moral responsibility is distributed between humans and artefacts? How can humans share responsibility with artefacts, given that moral responsibility traditionally entails moral blame and praise? Is it meaningful to blame artefacts for the harm they cause? Furthermore, moral blame is often followed by deserts. Does that mean artefacts should be punished as well? Hanson says, "We may very well blame them and blaming, no less than punishing, is generally taken to be a marker of moral responsibility" (Hanson 2009).

Nevertheless, Hanson cautions against using deserts as the sole criterion for attributing moral responsibility, emphasizing that considerations of deserts typically arise subsequent to the determination of responsibility. While I agree that deserts should not be the sole criterion, Hanson does not elaborate on

how we should conceptualize the shared responsibility between artefacts and human users. What aspects of responsibility pertain to artefacts, and what are the practical implications of attributing moral responsibility to extended agency? Addressing these questions is crucial for understanding the concept of joint responsibility. I do not think it is a short step from agency to moral responsibility, as Hanson suggests. It is plausible to think of machine actions, such as those decisions made by the Eurotransplant AI, as performed by humans and machines, and moreover, recognise the moral significance of these machines in shaping human behaviours. However, joint responsibility doesn't necessarily follow from this. Our agency could extend to other objects, yet not our moral responsibility. The goal seems to be that we want to acknowledge the moral role of autonomous machines but also do not want to attribute mental capacities and moral responsibility to artefacts. In the next section, I will talk about Deborah Johnson and Thomas Powers' (2008) alternative surrogate agent model to think about the moral agency of computer systems.

4. Surrogate Agency Model

Johnson and Powers (2008) argue that the human surrogate agency is a good model for understanding the moral agency of computers. Their argument centres on the premise that “Computer systems, like human surrogate agents, perform tasks on behalf of users. They implement actions in pursuit of the interests of others. As a user interacts with a computer system, the system achieves some of the user’s ends” (Johnson and Powers, 2008). Examples of human surrogate agents they provide include lawyers, tax accountants, and estate executors. What is common to them is that they mainly consider what the client wants, not what they want. A surrogate agent adopts a third-person perspective and pursues the second-order interests or desires of a client. I believe this perspective is crucial for reevaluating the “autonomy” of machines that contribute to the responsibility gap.

Evans, Robbins, and Bryson (2023) draw attention to the problematic use of anthropomorphic vocabulary—such as “autonomous”—when describing these technologies. To gain a better understanding of the autonomy certain machines possess, they refer to Virginia Dignum’s definition of task autonomy:

“...the ability of the system to adjust its behavior, by forming new plans to fulfill a goal, or by choosing between goals.” The machines in question have autonomy of means, but not autonomy of goals. Achieving the latter requires the ability to “...introduce new goals, modify existing goals, and disengage from active goals” (Dignum, 2019). Building on this concept, Evans, Robbins, and Bryson argue against the popular view that humans and machines can collaborate. For collaboration, the capacity to pursue ends of one’s own choosing is needed; current AI systems are not even close to having this kind of capacity. They offer the following perspective on ChatGPT:

Much like fictional genies, there is nothing these systems seek to achieve beyond the wishes (or commands) of their users. Their agentive capacity is limited to responses to well-named 'prompts'. Much like oracles, generative AI can offer little to no causal explanation for its outputs. These models are simply capable subsidiary agents, greasing the wheels of human action and collaboration. Thus, collaboration is simply not possible for such machines and products. This likely applies to other current human–AI interaction buzzwords that require similar capacities—like 'teammate' or 'partner.' What, then, is happening when humans and machines are seen as working together?

As an alternative, they propose non-collaborative joint action as a way of conceiving human-machine interaction: humans and machines can jointly work toward a common end. Similar to Hanson's perspective, proponents of the surrogate agency model believe that moral agency could be shared between a human and a machine, yet unlike Hanson, they do not argue for shared moral responsibility. I think this is an important stance to maintain. For example, Johnson and Power claims that computer systems possess a moral agency—distinct from that of humans— yet they don't see it necessary to make the inference computer systems can be morally responsible. They note, “It is plausible that the moral agency of computer systems does not entail responsibility, liability, and blame” (Johnson and Power, 2008).

The parallel drawn between human surrogate agents and current AI-powered machines holds appeal in various ways. Unlike instrumentalism 2.0, it acknowledges the increasing autonomy of machines. Unlike machine ethics, it refrains from over-attributing moral capacities to machines, while recognizing

the moral significance machines hold in shaping human behaviors. Although Power and Johnson do not claim that moral responsibility necessarily follows from moral agency, they also do not argue against it. I contend that we should resist the idea that machines could be morally responsible for two primary reasons. The first reason stems from the standard account of responsibility attribution, which requires an autonomous agent with relevant mental states, such as intentions, desires, and beliefs. The autonomy of the agent is intricately linked to their mental capacities. While machines may possess autonomy of means, this alone is insufficient for attributing moral responsibility. The second reason has to do with the possibility of letting the responsible human agents off the moral hook. Holding machines morally responsible could easily lead to diminished accountability for human agents who are really responsible.

The disanalogy between human surrogate agents and autonomous machines in terms of responsibility attribution could be a reason for questioning the model. Human surrogate agents could easily be held responsible for their actions. When a tax accountant makes a mistake on behalf of their client, we do not typically hold the client responsible. I think we need a model that could accommodate both joint action and moral responsibility of human agents. Consider the following example as a starting point for constructing a better model: Human truffle hunters employ pigs to locate truffles. We can readily conceive of truffle hunting as a joint action carried out by both the pig and the human. The pig and the human collaborate to achieve a shared goal, with the human determining the collective objective. Furthermore, the pig possesses autonomy over its means. In this regard, it resembles a surrogate agent; the human hunter delegates some of their tasks to the pig, relinquishing control over the pig's actions in pursuit of this goal. Additionally, it is well-known that pigs can cause significant damage to the terrain. Let's imagine a scenario where, during a truffle hunt, the pig enters someone's garden and causes substantial damage. In such a case, who would we hold responsible for the pig's behaviour? It would typically be the owner of the pig who is held accountable. From this perspective, machines seem to bear more resemblance to pigs than tax accountants in terms of responsibility attribution.

5. Conclusion

I started with the standard account of responsibility attribution, focusing on the control and knowledge conditions. Then, I explained how machine learning algorithms present a challenge to this framework, leading to the emergence of the responsibility gap. The increasing autonomy of machines undermines the control condition. As a potential solution, I delved into Hanson's joint responsibility approach, which advocates for the distribution of moral responsibility across a network of both human and machine components. While I agreed with Hanson on the idea that the agency can be shared between human and machine components, I objected to the idea that joint responsibility follows from an extended agency. Instead, I proposed Johnson's and Power's surrogate agency model as a more suitable framework for conceptualizing certain human-machine interactions. However, I also pointed out the disanalogy between human surrogate agents and autonomous machines in terms of responsibility attribution. There are compelling reasons for not attributing moral responsibility to machines, and I think we need a model that could accommodate this.

References

- Dignum, Virginia (2019) *Responsible artificial intelligence: how to develop and use AI in a responsible way* (Vol. 2156).
- Evans, Katie D., Robbins, Scott A. and Bryson, Joanna J. (2023) Do We Collaborate With What We Design?, *Top. Cogn. Sci.*.
- Gunkel, David J. (2020) Mind the gap: responsible robotics and the problem of responsibility, *Ethics and Information Technology* 22 (4), 307-320.
- Hanson, F. Allan (2009) Beyond the skin bag: On the moral responsibility of extended agencies, *Ethics and Information Technology*, 11(1), 91–99.
- Johnson, Deborah G., & Powers, Thomas M. (2008) "Computers as surrogate agents", in: J. van den Hoven, & J. Weckert (Eds.), *Information technology and moral philosophy*, 251–269. Cambridge: Cambridge University Press.
- Matthias, Andreas (2004) "The responsibility gap: Ascribing responsibility for the actions of learning automata", *Ethics and Information Technology*, 6 (3), 175–183

The Exportability of Embedded Social Kinds

Denis Kazankov (Vienna, Austria)

Abstract

This paper concerns what I call *embedded social kinds*. These are the social kinds that seem to have different grounding conditions for their instantiation across different societies. What is puzzling about these kinds is that their embeddedness seems to hinder the exportability of their grounding conditions to the societies that don't accept them, even in the cases where such exportation seems plausible. For example, most people would agree that Genghis Khan committed atrocities that made him a war criminal already during his life even though the international treaties which legally established the kind *war criminal* didn't exist back then. In this paper, I explore the possibility of developing a plausible interpretation of embedded social kinds that allows for their exportability. I first discuss Schaffer's proposed interpretation (2019) which treats embedded social kinds as relational properties. I argue that this interpretation is unsatisfactory because it fails to reflect two important dimensions along which the exportability of embedded social kinds varies: (i) the extent to which these kinds are morally significant, and (ii) normative differences between their alternative grounding conditions. I then present my alternative interpretation of embedded social kinds that does justice to the two dimensions. On this interpretation, grounding conditions for instantiating exportable embedded social kinds come into effect through the mechanism that ensures that they are based upon the correct moral attitudes. I conclude by briefly addressing some foreseeable objections.

1. Introduction

This paper concerns what I call *embedded social kinds* (hereafter *ESK*). ESK can be characterised in ground-theoretic terms as the social kinds that seem to have different grounding conditions for their instantiation across different societies in which they are embedded. Grounding is understood here as a non-causal explanatory relation between facts that 'links the world across levels' by producing less fundamental facts from more fundamental facts (Schaffer 2012: 122). Accordingly, the facts about what objects instantiate ESK seem metaphysically explained by distinct sorts of more fundamental facts across different societies.

To illustrate, consider the social kind *political leader*. In some societies, it might be sufficient to be formally selected for a particular public office to count as a political leader. Other societies might require that, besides holding a formal status, a political leader is engaged in specific public activities. Moreover, what

public engagement is required might also vary across different societies. While in earlier societies, in-person public engagement might have been sufficient, in contemporary societies, being active on social media might also be required.

Similar cases can be made about many other social kinds such as *social elite*, *war criminal*, *expert*, *religious organisation*, *legal gender*, *permanent resident*, *tax resident*, *university*, *juridical person*, *marriage*, *money*, or *private property*. Insofar as the realization relation is a species of grounding, many of these kinds can be described as what Griffith (2018a: 397; 2018b) calls *multiply realizable social kinds*. However, the variability of grounding conditions for instantiating ESK is meant here to cover more than multiple realizability. What can make a social kind embedded is not only that it is realized differently in different societies, but also that when embedded in some societies, it might be impossible to instantiate it because it isn't assigned any grounding conditions there. For instance, even if the kind *political leader* wasn't multiply realizable, as long as there are some anarchist societies that lack any grounding conditions for its instantiation due to categorically rejecting political authority, it would still count as ESK.

What is puzzling about ESK is that their embeddedness seems to hinder the exportability of their grounding conditions to the societies that don't accept these conditions, even in the cases where such exportation seems plausible. To use Epstein's example (2015: 124), most people would agree that Genghis Khan committed atrocities that made him a war criminal already during his life even though the international legal treaties such as the Nuremberg Principles, Geneva Conventions or The Rome Statute of ICC – which legally established the kind *war criminal* as we know it today and thus fixed its grounding conditions – didn't exist back then.

However, it seems difficult to reconcile our intuitions about the exportability of ESK with their nature. After all, if grounding conditions for being ESK vary across different societies, it seems to follow that what conditions objects must satisfy to instantiate these kinds is constrained by what societies the objects are situated in. That is, ESK suggest that even if someone meets the conditions for being a war criminal, as they have been established by current international law, this person still cannot be a war criminal as long as they are situated in a past or hypothetical society which either doesn't accept any

conditions for being a war criminal or accepts their own but different conditions for being a war criminal. This then brings us to the unpalatable conclusion that there were no war criminals throughout most of human history until international law invented them after WWII.

In this paper, I explore whether it's possible to work out a plausible interpretation of ESK that allows for their exportability. Here is how the paper unfolds. Firstly, I discuss Schaffer's proposed interpretation (2019) which treats ESK as relational properties. Secondly, I argue that Schaffer's interpretation is unsatisfactory because it fails to reflect two important dimensions along which the exportability of ESK varies. Thirdly, I present an alternative interpretation of ESK that involves a reconsideration of the mechanism through which grounding conditions for instantiating ESK come into effect. I conclude by briefly addressing some foreseeable doubts about my argument.

2. The Relationalist Interpretation of Social Kinds

Relationalism about social kinds, as Schaffer (2019: 763-765) introduces it, is the view that there are no monadic social kinds *simpliciter* but only relational social properties involving multiadic relations between individuals and social rules. The relationalist analysis of the kind *political leader* looks as follows: we have first a relation *X being a political leader by Y* consisting of a placeholder X for the first relatum, which is an object fulfilling the role of political leader, and the placeholder Y for the second relatum, which is social rules stating the conditions for instantiating the relation. Once we fill the second placeholder Y with specific rules, we arrive at a relational property such as *being a political leader given by the rules of Ukraine* or *being a political leader by the rules of the US*.

Given relationalism, the exportability of ESK can be secured. To understand why this is so, we must appreciate that social rules are posited here merely as relata. As such, they state grounding conditions for instantiating the relational property in which they feature but they don't impose any constraint upon the scope of application of these conditions. This is due to the fact that relations can arguably be cross-spatial, cross-temporal, cross-world and so, as a relatum, social rules don't have to be situated in the same world, time or place as the

other relata of the relation they feature in. Also, even the grounds of the facts about such relations don't have to obtain in the same world, time or space as the objects they apply to are situated in. So social rules can apply to any object that meets the conditions they fix.

With the rules stating grounding conditions for instantiating ESK having an unconstrained scope of application, we can then say things like that even though Genghis Khan didn't instantiate the kind *war criminal by the conventions of his own society*, what matters is that he instantiated the kind *war criminal by Geneva Conventions* simply in virtue of satisfying their content irrespective of what society he was situated in. This interpretation then allows us to export ESK, created in one society, to any other society.

3. Two Problems with the Relationalist Interpretation

My issue with Schaffer's relationalist interpretation of ESK is that it fails to reflect that their exportability varies along two dimensions: (i) the extent to which ESK are morally significant, and (ii) normative differences between their alternative grounding conditions.

Let me start with explaining the first dimension. Social kinds seem to differ in how difficult it is for us to think about them from a morally neutral standpoint. On the one hand, there seems to be a category of social kinds that are so closely tied up with our moral attitudes that adopting a value-free perspective towards them seems impossible. I call these kinds morally significant social kinds. Some examples of them include *war criminal, genocide, rape, mafia, cult, racist, treason, human rights, civic society, dissident, social mobility, community engagement, or refugee*. One heuristic method for testing whether a social kind K is morally significant is to ask: is there any thin or thick moral description which is such that it sounds contradictory (or almost contradictory) to say that the description doesn't apply to the instances of K or to the state of instantiating K? K is morally significant when the answer is 'yes'. For example, it sounds contradictory to say that it isn't morally bad if someone is a war criminal. Or else, it sounds contradictory to say that leading the life of a dissident isn't morally admirable.

On the other hand, there are also kinds that seem much less morally significant for us such as *social elite*, *public transport*, *industry*, *office worker*, *permanent resident*, or *legal contract*. This is evidenced by the observation that these kinds seem to fail the above test, as it's difficult to think of a moral description which is such that it sounds contradictory to deny that it applies to them. Of course, this isn't to say that we never associate these kinds with any moral attitudes, but only that it isn't part of their nature to call for moral evaluation.

The distinction between morally significant and morally insignificant social kinds certainly involves a lot of grey area. Social reality is complex, and thus most distinctions through which we carve it out include borderline cases. But this doesn't show that the distinction isn't real, provided that we can find enough kinds that neatly fall into one of the two categories. In our current debate, the distinction seems especially relevant because our inclinations to export morally significant ESK seem much stronger than our inclinations to export the other ESK.

The aforementioned example of *war criminal* will bring this point home. We find it very challenging to accept that there were no war criminals in past societies. One potential explanation for this response is that we consider the existence of the kind *war criminal* in our society to reflect a moral attitude that we consider to be universally correct across all societies. This attitude amounts to our general disapproval of any war atrocities coupled with the conviction that it would be deeply unjust towards their victims if their perpetrators weren't treated as war criminals by their societies.

What supports this explanation is that it's considerably easier for us to refrain from exporting those ESK that are less morally significant for us, such as the above-listed ones. That is, it doesn't seem equally challenging to accept that these kinds might have very different grounding conditions in the societies that are structured very differently from our society. Doing so might even be vital for understanding these societies. For instance, there might be a society in which the distribution of power-related resources and its members' attitude towards these resources are extremely different to our society. Analysing the social elite of this society through the grounding conditions for being a social elite in our society might blind us to what makes its organisation unique.

Hence, there is a clear difference between our inclinations towards exporting morally significant ESK as opposed to morally insignificant ESK. This difference cannot be explained if we treat all ESK as exportable in virtue of their relational structure.

Additionally, there is another dimension which bears on the exportability of ESK but which relationalism disregards. Grounding conditions that even a morally significant ESK has in one society aren't always exportable to other societies. Rather, they are only exportable to the extent that they are morally preferable to their alternatives. To illustrate, imagine two societies, S1 and S2. S1 accepts the legislation that doesn't classify individuals as war criminals even if they commit serious atrocities during wartime. By contrast, S2 accepts the legislation that classifies people as war criminals in accordance with the current international law. The exportability of grounding conditions for being a war criminal doesn't appear symmetric between S1 and S2. While S2 is justified to classify people in S1 as war criminals according to their accepted own grounding conditions, the reverse doesn't seem to obtain. This asymmetric judgement is substantiated by the fact that S2's grounding conditions for being a war criminal are morally better than S1's own conditions. Yet, the relationalist interpretation cannot account for this because it implies that all ESK are equally exportable. This is undesirable because exportability then cannot serve as a resource to which various societies can appeal when justifying their treatment of ESK in other societies according to their own standards while denying the reverse treatment from the other side. Therefore, we shouldn't content ourselves with the relationalist interpretation of ESK.

4. An Alternative Interpretation

I will now propose an alternative interpretation of ESK which does justice to the aforementioned two dimensions of exportability. What I suggest is to reconsider the mechanism through which grounding conditions for being a morally significant ESK come into effect so that we allow them to be asymmetrically exportable.

My interpretation is guided by two key observations. Firstly, given the close connection between morally significant ESK and a society's moral attitudes, it's

natural to think that these attitudes are reflected both in grounding conditions for their instantiation as well as in the society's background arrangements fixing them. Secondly, it appears to be in the vital interest of all societies with exportation inclinations to include such morally significant ESK whose grounding conditions reflect correct moral attitudes. As argued above, a society's inclinations to export their grounding conditions for instantiating ESK to other societies seem motivated by the belief that these conditions are morally preferable. It would be glaringly incoherent for a society to have such inclinations if it didn't care about having correct moral attitudes.

Accordingly, an attractive way for a society to ensure that its morally significant ESK are based on correct moral attitudes is to condition what grounding conditions for their instantiation come into effect by the facts about whether their own background arrangements manifest the correct moral attitudes. If they are correct, the grounding conditions recommended by these arrangements get fixed. However, if they are incorrect, the grounding conditions recommended by alternative background arrangements manifesting the correct moral attitudes get fixed instead. That is, the society employing this strategy sets up a bifurcating mechanism for fixing the grounding conditions of its morally significant ESK whose outputs depend on what conditions are morally preferable.

Let me illustrate the idea by a real-life example. Before 1991, there was a legal agreement in England that the non-consensual sexual intercourse within marriage is immune from criminal liability and thus doesn't qualify as the crime of rape, legally speaking. This agreement rested upon the confused assumption that the contract of marriage includes husband's 'right to sex', which the wife recognises when entering the contract and cannot retract it later at any point of their marriage. Yet, when the marital rape exemption was abolished in 1991 in England, the leading judge appointed by the House of Lords, Lord Keith of Kinkel stated the rape exemption was a common law fiction which had never been a true rule of English law (ECHR: 1995). Assuming that he was correct, it follows that non-consensual sexual intercourse within marriage constituted the crime of rape even in pre-1991 England, despite not being legally treated like that for most of England's modern history. This seems like a paradigmatic case of the justified retroactive exportation of grounding conditions for instantiating an ESK.

My proposed interpretation can explain why this exportation is justified. Insofar as the pre-1991 English society cared about having the legal kind *the crime of rape* fixed by the correct moral attitudes about non-consensual sex, the society can be interpreted as having set up such a fixing mechanism for what counts as the crime of rape that didn't include only its actual legal arrangements, but also an implicitly accepted activating rule. This rule stated that the society's actual legal arrangements become activated to fix grounding conditions for being the crime of rape if they manifest the correct moral attitudes about non-consensual sex, but also that if they don't do so, then the society defers instead to alternative legal arrangements manifesting the correct moral attitudes.

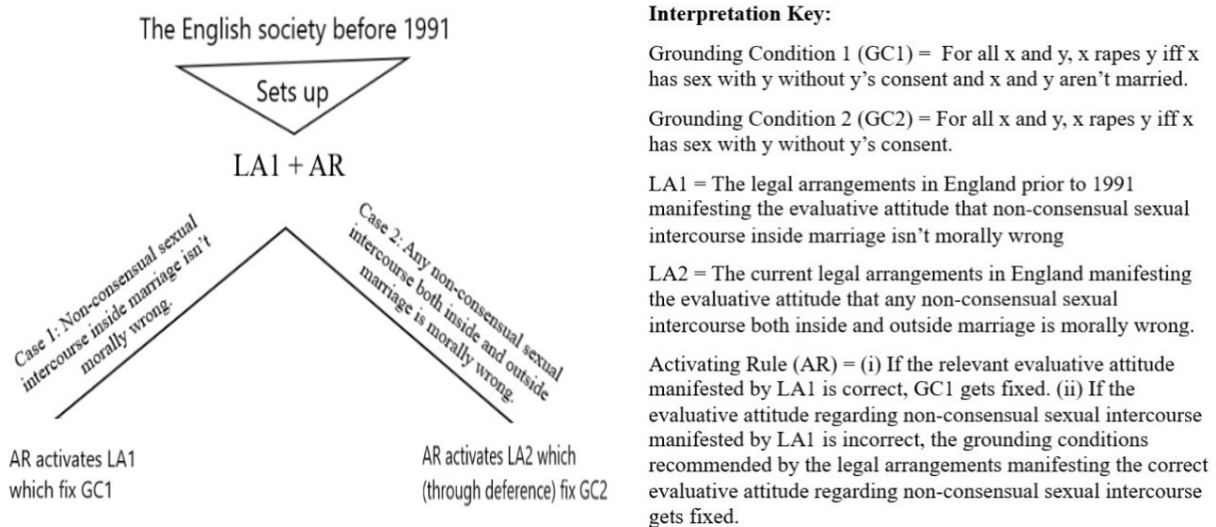


Figure 1: The mechanism for fixing grounding conditions for instantiating the crime of rape in pre-1991 English society.

As Figure 1 shows, the pre-1991 English legal arrangements manifested moral attitudes that didn't consider non-consensual sex inside marriage to be morally wrong. By contrast, the current English legal arrangements manifest moral attitudes that consider it to be morally wrong. It's obvious that the current attitudes are the correct ones; thus, Case 2 in the diagram is true. On my interpretation, the pre-1991 English society should be interpreted as deferring, by means of the activating rule, to its current legal arrangements in fixing what grounds that something is the crime of rape. This explains why the post-1991 English society is justified in exporting the grounding conditions for being the crime of rape recommended by their current legal arrangements to the pre-1991 English society, but not vice versa.

5. Conclusion

My proposed interpretation is both descriptive as well as ameliorative. Reasonable societies can be interpreted as adopting my proposed mechanism for fixing grounding conditions for instantiating morally significant ESK. But even if they turn out not to adopt it, they are at least strongly recommended to do so if they want these kinds to be based on correct moral attitudes.

Let me conclude by briefly addressing three doubts that the reader might have about my argument. Firstly, due to the limited scope of my discussion, I couldn't address Epstein's treatment of exportability, according to which social kinds are exportable because of having their grounding conditions fixed through the metaphysical relation he calls *anchoring* that enables them to be instantiated in any context (2015: 123-124, 2019). Nevertheless, since Epstein's theory presents exportability as a uniform feature of social kinds (2015: 69), I suspect that the challenge I raised against relationalism applies to it as well. Secondly, my proposed interpretation of ESK is compatible with legal positivism, i.e. the view that the conditions of legal validity are determined by social facts. In particular, it's compatible with the inclusive legal positivism according to which the source of legal validity is always social even if a society sometimes makes it conditional upon moral considerations (Waluchow 1994, Krammer 1999). Thirdly, it's true that since I talked about moral attitudes as being subject to the standards of correctness that are stable across different societies, my proposal might sound controversial to moral relativists. This is, however, harmless to my argument because moral relativists aren't my target audience in the first place. My argument is designed to convince the readers who are inclined to export grounding conditions for instantiating morally significant ESK from one society to other societies. If I'm right that their inclination stems from the assumption that exportable morally significant ESK manifest universally correct moral attitudes, then moral relativism is a non-starter for these readers.

References

Epstein, Brian. (2015) *The Ant Trap: Rebuilding the Foundations of the Social Sciences*, Oxford: Oxford University Press.

Epstein, Brian. (2019) "Anchoring versus Grounding: Reply to Schaffer", *Philosophy and Phenomenological Research* 99, 768-789.

European Court of Human Rights [ECHR]. (1995) CR v United Kingdom, Decision on merits, App No 20190/92, A/355-C, IHRL 2595.

Griffith, Aaron M. (2018a) "Social Construction and Grounding", *Philosophy and Phenomenological Research* 97, 393-409.

Griffith, Aaron M. (2018b) "Social construction: big-G grounding, small-g realization", *Philosophical Studies* 175, 241-260.

Krammer, Matthew H. (1999) *In Defense of Legal Positivism: Law Without Trimmings*, Oxford: Clarendon Press.

Schaffer, Jonathan (2012) "Grounding, transitivity, and contrastivity", in: Fabrice Correia and Benjamin Schnieder (eds.) *Metaphysical grounding: Understanding the structure of reality*, Cambridge: Cambridge University Press, 122-138.

Schaffer, Jonathan (2019) "Anchoring as grounding: On Epstein's the ant trap", *Philosophy and Phenomenological Research* 99, 749-767.

Waluchow, Will J. (1994) *Inclusive Legal Positivism*, Oxford: Clarendon Press.

Dummett and the Reality of Pain

Gilad Kenan (Tel Aviv, Israel)

Abstract

How can we account for the reality of other people's pain? And is it possible to doubt my own pain? Michael Dummett considers Wittgenstein's Private Language Argument as an incontrovertible and definitive argument for anti-realism with regards to pain, an approach which successfully circumvents the pitfalls both of realism and of behaviourism.

Dummett's work offers a reconceptualization of the disagreement between realism and different types of anti-realism. The paper begins with a general overview of his theoretical framework. It proceeds to an examination of his interpretation of Wittgenstein's argument in these terms. I show that according to Dummett both the realist and the behaviourist share a common premise, namely, giving primacy to the first person perspective, a premise successfully refuted by Wittgenstein.

Against this, a communal and non-reductive approach is key to the kind of anti-realism which Dummett champions. To elucidate and bolster his claims I consider a recent case of maltreatment in a Yale fertility clinic, where patients' pain was ignored and unaddressed. Their pain was considered unreal or exaggerated not only by the medical staff but also to some extent by the patients themselves. This sheds light on the way recognition of pain by others is not subsequent or secondary to the phenomenon of pain, but rather constitutive to it. It thus points to a kind of anti-realism which is grounded in the community rather than in subjective sensations, which is helpful for understanding Dummett's philosophical approach in general.

Is pain real? Is other people's pain real? Is it real just as much as mine? Is women's pain as real as men's? And is it possible to doubt my own pain? In what follows I discuss Michael Dummett's interpretation of Wittgenstein's Private Language Argument. As I hope to show, this discussion is in some respects exceptional on the backdrop of Dummett's own philosophical project, yet it illuminates it in important ways. By inspecting a recent incident in Yale's fertility clinic I argue with Dummett (and Dummett's Wittgenstein) against both realism and behaviourism. The discussion sheds light on the way recognition of pain by others is not subsequent or secondary to pain itself, but rather constitutive to it. It thus points to a kind of anti-realism which is grounded in the community rather than in subjective sensations.

1. Realism and its alternatives

Let us start with a general overview of Dummett's conceptual framework. Dummett advocates the analysis of metaphysical disputes in linguistic terms and argues that this move exposes a common underlying structure. He writes: "... there is a range of traditional metaphysical disputes relating to very different subject matters but sharing a common form, the form, namely, of a conflict between a realist and an anti-realist view of some class of statements. When we prescind from the particular subject-matter, we see a striking analogy between the arguments and counter-arguments used in such disputes". (1978: p. xxx)

We traditionally conceive of realism as arguing for the existence of certain entities. Dummett suggests however to define realism linguistically, as explaining meaning in terms of truth-conditions which transcend our capabilities to confirm them. The realist considers reality as fully determinate and every proposition as necessarily either true or false, irrespective of our ability to know or prove it, which amounts to acceptance of the Principle of Bivalence and Law of Excluded Middle.

Realism thus posits truths lying forever beyond our recognition. This might lead one either to a form of dogmatic, naive belief in what cannot be verified, or to a form of scepticism, denying it can ever be known.

Dummett defines anti-realism as any approach objecting to realism, and this can take several forms. Dummett's own preferred approach is an alternative theory of meaning inspired by mathematical intuitionism, whereby the meaning of a statement has to do with our method of verifying or justifying it. Typically this involves rejection of Bivalence, since a demonstrated lack of justification for a statement does not amount to a justification of its negation and might leave a statement undetermined.

An alternative form of anti-realism is reductionism, which Dummett conceives as a form of translation from one class of statements to another class, supposedly more fundamental. We may decide statements of the reducible class by translating them to single statements or conjunctions thereof in the fundamental class. Reductionism is not necessarily concerned with presenting an actual, feasible translation manual of this sort, only with the possibility in principle of such a translation (1993: 240-241).

Reductionism might resemble Dummett's own justificationist approach, where it counters realism and offers to explain the meanings of statements of the reduced class in terms of statements that we can examine, verify or refute. However, as long as reductionism is still ultimately based on a truth-conditional theory of meaning it fails to be justificationist and thus genuinely anti-realist in Dummett's terms. It can and usually does preserve Bivalence by considering all untrue statements as false.

There is another way to characterise the dispute informally. As was said above, from an anti-realist point of view realism is problematic since it implies that there is a whole range of true facts which will remain forever unknown, at least to us limited mortal beings. It is in this sense that Dummett sometimes characterises realism as entailing the view that if we were endowed with stronger epistemic capabilities we would have been able to know these truths. The realist is accused of implicitly presupposing a hypothetical superhuman perspective, which is able to inspect reality from above as it were. Contrarily, the anti-realist is committed to the limitations of human perception and develops their full consequences. However, as we shall see, this importantly does not amount to subjectivist idealism.

2. Pain: Realism and Behaviourism

Now let us turn to Dummett's discussion of pain. Dummett considers Wittgenstein's Private Language Argument as a solid case for anti-realism. He states his "conviction" that the argument "is incontrovertible", that the untenability of realism on this matter has thus "been definitively demonstrated" (1978:xxxiii). Such confidence and a sense of finality are exceptional in his writing and are never used in reference to his own arguments.

Nevertheless, Dummett does not dedicate much writing to the argument. To the best of my knowledge, he does so three times, rather briefly and in passing: once in his 1982 paper "Realism" (1993:237-239), another time in his book *The Nature and Future of Philosophy* (2010:140-144), and in a little more detail in the preface to his book *Truth and Other Enigmas* (1978:xxxii-xxxviii), where he brings it up to counter comments made by P. F. Strawson.

These discussions make different points. What is common is that they categorise Wittgenstein's argument as a special approach, different from Dummett's own usual reasoning and argumentation for anti-realism. Apparently then, we have here an exceptionally strong argument for anti-realism, unique in character and yet rarely mentioned and scantily analysed in Dummett's work. All this makes it interesting and deserving of observation.

Let us consider ascriptions of pain in the second or third person, as in the statement "Martha is in pain". Dummett's classification is helpful in categorising the different philosophical positions one may take. Realism about pain amounts to analysing this proposition in terms of truth conditions: it is true whenever Martha feels what I feel when I am in pain. The condition is a subjective sensation available to Martha herself and her alone. Although this statement in the third person is to be taken as determinately either true or false, it cannot be asserted with certainty. We might accept it as "simply true" or "simply false" without justification, and risk dogmatism and naivete, or rather maintain that it always lies beyond our reach, and risk some sort of scepticism.

Nonetheless, we do ascribe pain to others and consider such ascriptions true or false, and consider ourselves justified in doing so. Behaviourism is a reductionist approach to counter realism and to address the problems arising from it. It translates pain ascriptions to visible criteria: things like bleeding and crying, to be sure, but we can also extend it to verbal behaviour, surrounding evidence, preceding events, etc. The behaviourist might consider the assertion "Martha is in pain" as equivalent to a conjunction of these criteria, to a weighted majority or perhaps something more sophisticated. The important point for our discussion is that this method eliminates the room for a verification-transcendent truth, and thus for dogmatic realism or for scepticism. It also preserves bivalence, since if there is no grounds for the ascription of pain such a statement can be regarded as false.

Strawson finds behaviourism highly unconvincing. In the text Dummett alludes to, Strawson mentions one implication that might be taken as a reason for rejecting it: an asymmetry, a divorce even, between "I am in pain" and "you are in pain". Under behaviourism ascription of pain would have a different meaning in the first person, compared with second and third person

ascriptions. Against this Strawson says that “our general theory of the world” is just that we ascribe to others the same sensations which we feel, and this must entail accepting “justification transcendent” statements as true. Clear realism, in Dummett’s terms.

Dummett himself reminds us of another challenge to behaviourism: the possibility of deceit or other forms of misrecognition. One might pretend to be in pain and stage the whole thing, or conversely, endure pain stoically without any outward expression. *Prima facie*, these possibilities introduce a cleavage between subjective feeling and publicly available signs, for which behaviourism has to account.

3. Dummett’s Wittgenstein

For Dummett these are serious challenges, yet he rejects both answers. Strawson’s mistake in his eyes is to consider them as the only ones possible. Assisted by Wittgenstein, he identifies a common assumption to realism and behaviourism which circumscribes the dispute in advance. He writes:

The realist and the behaviourist disagree about what makes an ascription of pain to another true: the realist says it is his having what I have when I am in pain, the behaviourist that it is his exhibiting pain-behaviour. Tacitly, they both believe in the private ostensive definition: at least in his heart of hearts, the behaviourist does not think, any more than the realist does, that there is any problem about what makes ‘I am in pain’ true, when said by me. (1978: p. xxxvii)

Both sides accept first-person reflexive ascriptions of pain as primary and non-problematic, and the problem is thus framed as how to bring ascription to others to meet this standard. For Dummett the merit of Wittgenstein’s argument is exactly that it rejects this premise and framing altogether. As adults, recognition of pain in others seems secondary to our own inward reflection, but in the process of acquiring language as children it is clear that externally identifiable signs of pain play an essential role. It is necessarily this shared public domain that grounds the functioning of language.

Dummett's approach, following his understanding of Wittgenstein, is therefore justificationist but not reductivist. It does not try to eliminate the subjective sensations out of the explanation, but to insist that these are not the primary or ultimate justifications for ascriptions of pain. He writes:

The account of meaning in terms of truth-conditions has to be replaced by one in terms of the conditions under which we are justified in making such statements, including ones when the justification may be overturned; and what justifies a statement of this kind does so only in view of the fact that certain general connections hold. (1978: p. xxxvii)

Two considerations accompany this argumentative move. First, language acquisition relies on a generally consistent connection between internal feeling and external signs. Deceit cannot be the rule, or deceit itself would not be possible.

Secondly, ascriptions of pain albeit many times true are never conclusive. We might retract our former ascription of pain when more data is available. This makes truth provisional and corrigible and introduces an irreducible gap between truth and certainty; but it amounts neither to scepticism nor to a position whereby these pain ascriptions are neither true nor false. Dummett writes that "it is essential for the concept of pain both that we judge on the basis of behaviour and that such judgments are frequently defeasible." (p. xxxviii)

But is this position anti-realist? It is indeed quite different from the usual course of argument Dummett advances. It does explain meaning in terms of justification or verification, relying on publicly available evidence. But it does not entail a rejection of Bivalence. If there is no evidence justifying such an ascription, it is taken as false. Either one is in pain or is not in pain.

Dummett seems to make here another significant compromise in accepting a rift between certainty and truth. It seems as though what the realist previously considered as a gap between justifiable ascription and ultimate truth is now mapped onto a gap between provisional truth and certainty, respectively. However, this is the case only if we retract to a position giving primacy to the first person. The point is to see how the external perspective is not secondary and subordinate to the internal subjective one.

4. The Yale Clinic Case

In order to clarify this, let us consider an example: a recent lamentable case which took place in a fertility clinic in Yale around 2020 and caught public attention in a podcast titled “The Retrievals”. As is customary, the clinic was conducting intrusive procedures with pain medication, fentanyl. Or so it was assumed until it was discovered that the fentanyl was regularly stolen by a nurse for personal use and replaced by saline solution. It turned out in retrospect that patients underwent these procedures in excruciating pain, torment really, totally sober.

It is not that they haven’t noticed. They cried and contorted. They expressed their pain and tried to query the medical staff if these sensations were normal and expected. But this was not taken seriously by the staff, since it was assumed by all that they were strongly medicated. Their first-person ascriptions of pain were ignored and things kept going as usual. It turns out that this is not uncommon, that there is a documented tendency of medical staff to disregard or undervalue women’s expressions of pain.

This tragic story shows us that in practice subjective sensations are not always taken as ultimate as we many times suppose them to be in philosophical discourse. Note that this case is different from deceit: it was never claimed that these women were lying, but rather that they are over-sensitive, that they exaggerate or convince themselves; that what they think they feel is unreal: “you didn't experience what you believe you experienced” (ep. 5). This does not second-guess the fact that this person does think they are suffering, but still doubts the objective reality of this sensation.

The eventual recognition that these patients were in fact in horrific pain was not due to a regained faith in their testimonies but to new external evidence: a loose cap led to investigation and finally exposed the theft. To this day it is still unclear how many of the patients underwent procedures without pain-killers. The problematic of recognition is repeated in court, where claims must be corroborated by external evidence, and lack thereof keeps the pain of many patients in an undetermined state.

What was said so far might be taken as congruent with a realist approach to pain. But what some patients describe goes a step further. They seem to have also dismissed their own experience in the face of perceived objective evidence to the contrary and the insistence of the medical staff. Patients say things like "I just felt like, nobody had heard what I was saying, so it just didn't exist"; or "I felt crazy [...] am I being difficult? you just question your sense of self." (ep.1)

Not only is it possible to prefer other forms of evidence over people's claims over themselves, but it is even possible for a person to accept this external judgement over and against their own sensations. Pace Wittgenstein who writes "it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself" (PI §246), we see that people do in fact doubt whether they are in pain, or at least the severity of it. Whereas what he says about identity conditions (§253-254) is relevant here: in order to measure pain we have to be able to compare it, therefore rely on publicly available information, and this goes also to our own sensations. It is not paradoxical, incoherent or nonsensical for a person to rely on external criteria to evaluate, mitigate or override internal sensations.

This means that, in Strawson's words, "our general theory of the world" does allow that first person subjective sensations are not primary, that they yield to objective, public evidence. This possibility derives from the fact that to begin with, recognition of pain by others is constituent to the phenomenon, rather than a secondary addition to an already self-conscious reflexive subject, whose sensations are immediate, evident and certain.

As has been stated above, overriding first-person ascriptions cannot be the rule or even the ordinary case. There is necessarily an overall harmony between internal sensations, external expressions of pain and recognition by others. But this does not preclude instances of contradiction between them or some indeterminacy. Furthermore, there is no reason to assume that ordinariness and exceptionality are evenly distributed in society. That certain classes of people are more prone to such second-guessing of sensations and inner judgements is not inconsistent with this prerequisite of general harmony.

5. Viewing Oneself from Outside

The shift in primacy from the first-person perspective to the perspective of the community is key not only for the analysis of pain but throughout Dummett's thought. I claim it is an essential feature which sets his position at a distance from other possible anti-realist views.

Across Dummett's writings the most prevalent pronoun is "we" rather than the usual philosophical "I". For him, justifications, knowledge and truth are governed and even constituted by shared communal norms. For thought and meanings to be objective they ought to be public and common. Thus it is not I nor you who judge whether P is true, it is we who verify, justify and judge true. And this is correct also when P is "I am in pain" or "you are in pain".

In a sense, we are equal in front of this collective perspective. There is a regained symmetry between first, second and third person ascriptions of pain. There is no precedence to one over the other. But in another important sense, we are not totally equal, as we have just seen, and some are treated differently from others.

I mentioned that realism might be characterised as the hypothetical possibility of occupying a superhuman point of view, looking over from above and safeguarding truth where it is transcendent to our limited human capacities. Does Dummett's thought submit here to a form of realism in this sense?

Dummett suggests, I claim, that as language users, we necessarily accept a perspective which is external to our own. It is not reducible neither to a simplistic and restrictive conception of external evidence nor to a sum of subjective perspectives. Yet it does not transcend them either. In order to account for reality we ought to recognize its intersubjective character, its dependence on recognition. This is no less true for something as intimate as pain.

References

- Dummett, Michael (1978) *Truth and other enigmas*. Harvard University Press.
- Dummett, Michael (1991) *The logical basis of metaphysics*. Harvard university press.
- Dummett, Michael (1993) *The seas of language*. Clarendon Press.
- Dummett, Michael (2010) *The nature and future of philosophy*. Columbia University Press.

Strawson, P. F. (1976) "Scruton and Wright on Anti-realism etc.", in: *Proceedings of the Aristotelian Society* (Vol. 77, pp. 15-21). Aristotelian Society, Wiley.

Wittgenstein, L. (1968) *Philosophical investigations*. G. E. M. Anscombe (ed.), Blackwell Ltd.

Acknowledgement

I wish to thank Anat Matar, Ittamar Erb and Uri Brun for their help.

Possibilism and the “Wrong Kind of Object” Problem

Tingjiang Kuang (Edinburgh, UK)

Abstract

Theories concerning fictional entities and fictional names face the so-called “wrong kind of object” problem. The problem, as outlined by Semejin and Zalta (2021), is that if fictional names denote abstract objects, then some parafictional statements (e.g., Sherlock Holmes is a private detective) would involve the wrong kind of objects (properties only applicable to concrete objects are applied to the abstract objects); if fictional names denote concrete objects in fictional worlds or worlds of pretense, then some metafictional statements (e.g., Sherlock Holmes is a fictional character created by Conan Doyle) would involve the wrong kind of objects (properties only applicable to abstract objects are applied to the concrete objects). This essay addresses the problem from the perspective of possibilism about fictional entities and fictional names. A solution based on two-dimensionalism and counterpart theory will be given. It will be argued that a metafictional sentence is true if considered uttered in a fictional world but evaluated in the actual world. If the sentence is considered uttered in a fictional world, a fictional name picks out a concrete object, and a proposition about the concrete object is determined. This proposition is true if evaluated in the actual world because there is a fictional-person idea (Everett & Schroeder, 2015) in the actual world, which is the counterpart of the denoted concrete object. As a result, the “wrong kind of object” problem is circumvented, as it is the idea that properly receives ascriptions of metafictional properties.

Introduction

There are multiple kinds of fictional discourses that semantics has to address. For example, consider the following fictional statements:

- 1) Sherlock Holmes approached the body, and, kneeling down, examined it intently. (From *A Study in Scarlet*)
- 2) (In Conan Doyle’s stories), Sherlock Holmes is a private detective.
- 3) Sherlock Holmes is a fictional character created by Conan Doyle.

According to Recanati (2018), these statements can be categorized as fictional, parafictional, and metafictional. Statement 1) is a fictional statement that is included in Conan Doyle’s novel. It is used by Conan Doyle to tell a fictional story. Statement 2) is a parafictional statement that is meant to describe what is the case in the story. The statement is explicitly or implicitly prefixed by an “in/according to” fiction operator. Statement 3) is a metafictional statement that purports to attribute properties to a fictional individual outside the scope of fictional stories. According to Conan Doyle’s stories, Sherlock Holmes is a

flesh-and-bone human being, not a fictional character created by Conan Doyle. So, statement 3) is not about what is the case in fiction but what is the case from a metafictional perspective.

“Sherlock Holmes” appears in all the above three kinds of statements. A question is whether all occurrences of “Sherlock Holmes” have the same meaning. Semejin and Zalta (2021) argue that, at least, a fictional name should have the same meaning in parafictional and metafictional statements since there could be anaphoric links between the two kinds of statements. I will assume this is a proper approach to the question and focus on how to provide a unified explanation of statements 2) and 3).

Nevertheless, the remaining problem is that it is hard to give a unified interpretation of “Sherlock Holmes” in 2) and 3). Semejin and Zalta (2021) point out that if “Sherlock Holmes” denotes an abstract object (e.g., Zalta, 1988; Kripke, 2013), statement 2) becomes a category mistake, akin to claiming a number can be a private detective. Conversely, if “Sherlock Holmes” denotes a flesh-and-bone human in some possible worlds or worlds of pretense (e.g., Lewis, 1978; Walton, 1990; Everett, 2013), statement 3) becomes a category mistake, as no human can be a fictional creation. This dilemma is termed the “wrong kind of object” problem (Klauk, 2014).

In this paper, I will address the problem from the perspective of possibilism (Kroon & Voltolini, 2023, sec. 1.1). Possibilism claims that “Sherlock Holmes” denotes flesh-and-bone concrete objects in possible worlds rather than denoting anything in the actual world. The problem for possibilism is whether 3) is a category mistake, given that “Sherlock Holmes” denotes a concrete object in some possible worlds. I will show that there is a way to hold the thesis of possibilism without engaging in any category mistake.

My argument will go in the following way. In section 1, I will restate the “wrong kind of object” problem and explain why possibilism should care about it. In section 2, I will present an interpretation within a two-dimensional framework, which makes a concrete object in other possible worlds a fictional creation by someone in our world based on its counterpart relation to a

fictional-person idea. The interpretation leads to a two-dimensional framework of possibilism. In the end, I will respond to a potential worry about my solution.

1. The “Wrong Kind of Object” Problem

Not every predicate can be applied to abstract objects. For example, predicates involving spatial location cannot be used to describe an abstract object:

4) Number 4 lives at 221B Baker Street.

4) is a category mistake. It sounds totally nonsense. If anyone utters sentence 4) without any intention of metaphoric use, it is rational to doubt if the utterer really understands the words involved in 4). For, it is conceptually impossible for 4) to be the case. Similarly, applying some predicates to concrete objects can also cause category mistakes. Like,

5) Caesar equals to $2+2$.

This is a kind of mistake that we hardly make, supposing that we have an understanding of the words involved in 5). It is hard to see why anyone would make a mistake like 5) if he or she has no misunderstanding of the meaning.

The problem of the wrong kind of object says our fictional discourses involve this kind of mistake, assuming some theories of fictional characters. Semejin and Zalta (2021) claim that almost every theory of fictional characters also faces the problem.

Does possibilism, as Semejin and Zalta (2021) suggest, face the “wrong kind of object” problem? The key thesis of possibilism is that fictional names should denote possible objects in fictional worlds but not denote anything in the actual world. Like in our examples, “Sherlock Holmes” should denote a concrete object in some possible worlds. This works well with the interpretation of parafictional statements. Lewis (1978) has offered many analyses of parafictional statements, but I will only consider his analysis 1. Roughly speaking, the analysis says:

Para 1: A parafictional sentence s derived from fiction f is true iff s is true in every possible world where f is told as known facts instead of fiction.

In the following, I will call worlds where the related fiction is told as known facts fictional worlds. As an example of analysis, consider the parafictional statement we have seen:

2) (In Conan Doyle’s stories), Sherlock Holmes is a private detective

According to the above Lewisian analysis (1978), statement 2) is true iff Sherlock Holmes, as a concrete individual, is a private detective in each possible world where Conan Doyle’s stories are told as known facts rather than fiction. There are many controversies, but for our purpose, the Lewisian analysis of parafictional statements is already good enough.

However, possibilism has the danger of letting metafictional statements ascribe mismatch properties to concrete objects. No concrete object, as it appears, can be a fictional character. The interpretation of 3), according to possibilists, would be something similar to 5). It says a concrete person is invented by an author. But statement 3) does not sound like a category mistake. It is, at least, not conceptually wrong. Thus, the “wrong kind of object” problem seems to indicate that possibilism is wrong.

Some possibilists may think there is no need to care about the “wrong kind of object” problem because Lewisian analysis is not aimed to say anything about metafictional statements. Lewis (1978) himself is aware of metafictional statements but still decides to exclude them from his analysis. Why cannot possibilists say their theory is only about parafictional statements?

One reason for possibilists to be concerned about metafictional statements is that we can easily build anaphoric links between metafictional and parafictional statements (Recanati, 2018). Consider the following case:

6) Sherlock Holmes is a fictional character created by Conan Doyle. In Conan Doyle’s stories, he (Sherlock Holmes) is a private detective.

It is natural to think “he” in the parafictional part inherits the meaning of “Sherlock Holmes” in the metafictional part. That is to say, statement 6) first introduces a fictional name in a metafictional context and then brings it into a parafictional context. To interpret the parafictional part, possibilism has to explain what is the meaning of “Sherlock Holmes” in the metafictional part. Does it still denote concrete objects in possible worlds? Without answering

this, possibilism cannot even give a complete explanation of parafictional statements.

2. Fictional-Person Ideas As Counterparts

I have shown why the problem of the “wrong kind of object” matters to the possibilism of fictional entities. In this section, I will proceed to elucidate how possibilism can circumvent this problem. Firstly, I will introduce Everett and Schroeder's (2015) theory of fictional-person ideas and then illustrate how these ideas can aid possibilism. This will lead to the development of a two-dimensional framework for possibilism.

As Everett and Schroeder (2015) posit, ideas can be created by some people at some time. Einstein invented the idea of the theory of relativity in physics. The idea first came into Einstein's mind and then spread to the whole world. This feature indicates that ideas are mental states bearing contents.

Everett and Schroeder (2015) propose that ideas can be multi-located by having different parts in different minds, which entails a peculiar mereology of ideas. To make things simpler, I would like to introduce a type-token distinction instead of considering their part-whole relation. For instance, when talking about the idea of the theory of relativity, we are generally talking about an idea type. Einstein invented the idea in the sense that he created the first token of the idea type. Then, the idea type is spread through the world: different minds gradually acquire ideas belonging to this type.

The same thing happens to the fictional-person idea of Sherlock Holmes. Conan Doyle first invented the fictional-person idea of Sherlock Holmes, which means the first idea token of Sherlock Holmes was created in the mind of Conan Doyle. Thereafter, this idea permeated globally through his series of novels.

But what kind of idea is this? Everett and Schroeder (2015) claim a fictional-person idea of a fictional character is an idea for telling stories as if there was a particular person with particular properties. There is no Sherlock Holmes,

but we can talk as if there is such and such thing to others for entertainment or other aims. The idea of telling stories in this way was first created by Conan Doyle.

Everett and Schroeder (2015) argue that this idea is the perfect subject of properties in metafictional statements. As previously mentioned, the first-person idea of Sherlock Holmes was created by Conan Doyle, and this fact makes the statement “Sherlock Holmes was created by Conan Doyle” true.

Everett and Schroeder (2015) think that realists should identify fictional characters with fictional-person ideas, while irrealists can use fictional-person ideas to explain away metafictional statements that motivate realism. Artefactualism (Thomasson, 1999; Kripke, 2013) and pretense theory (Walton, 1990; Everett, 2013) are mainly considered by Everett and Schroeder. In the following, I will show that possibilism can also make use of fictional-person ideas to solve the problem of the “wrong kind of object.”

Possibilists, I contend, should consider statement 3) true of the fictional-person idea of Sherlock Holmes while maintaining that statement 2) is true of certain concrete objects in possible worlds. This will avoid the “wrong kind of object problem.” A fictional-person idea is suitable to be a fictional character created by Conan Doyle, and concrete objects in possible worlds are suitable to be a private detective.

To make statement 3) true of a fictional-person idea, it has to be evaluated in our world, the world where there is no Sherlock Holmes but only the fictional-person idea. Then, should possibilists say the fictional-person idea is denoted by “Sherlock Holmes” in our actual world? This suggestion is *prima facie* promising but contrasts the core thesis of possibilism. Possibilists want “Sherlock Holmes” only to denote objects in other possible worlds.

The dilemma here is that “Sherlock Holmes” should not refer to a fictional-person idea in the actual world, but statement 3) has to be true of a fictional-person idea in the actual world. I think a solution to the dilemma is to consider counterpart relations between concrete objects denoted by “Sherlock Holmes” in possible worlds with the fictional-person idea of Sherlock Holmes in the actual world. “Sherlock Holmes” denotes possible concrete objects in 3), but when evaluating the truth value of 3), what is under consideration should be

the counterpart of concrete objects in the actual world, namely the fictional-person idea.

A two-dimensional framework can make this evaluation possible. There are already some applications of two-dimensionalism to fictional discourses (Murday, 2011; Voltolini, 2019), but I would like to form my framework in a bit different way. My suggested framework would be something Chalmers (2006, pp. 65-66) calls contextual two-dimensionalism. It includes both the context of utterance and the context of evaluation (Stalnaker, 1978; Chalmers, 2006, pp. 65-66). Within the framework, statement 3) can be seen as uttered in a fictional world but evaluated in the actual world.

The content of a sentence is partly determined by the world of utterance. If statement 3) is uttered in the actual world @, then nothing is picked out by "Sherlock Holmes," so the content of 3) can be seen as an incomplete proposition <..., a fictional character created by Conan Doyle>. On the other hand, if 3) is uttered in a fictional world *f* where Dr. Watson writes the stories of a real person *p* named Sherlock Holmes, who is a famous private detective that lives at 221B Baker Street, the content of 3) will be a complete proposition <*p*, a fictional character created by Conan Doyle>. After determining the content of a sentence in the world of utterance, we can then consider whether the determined content is true in the world of evaluation. Suppose <*p*, a fictional character created by Conan Doyle> is evaluated in *f*. Then, the proposition is false because *p* is not a fictional creation of Conan Doyle in *f*. However, if the proposition is evaluated in @, I think the proposition is true. A fictional-person idea of Sherlock Holmes is the counterpart of *p* in @, and this idea would make <*p*, a fictional character created by Conan Doyle> true in @. A simplified evaluation procedure can be defined as follows:

Eva: Proposition <*p*, F> is true in a world *w* iff the counterpart of *p* in *w* is F.

<*p*, a fictional character created by Conan Doyle> is true in @ due to the fact that there is an idea *i*, which is a fictional character created by Conan Doyle and is a counterpart of *p* in @. Following Lewis (1968), the counterpart of an object in a possible world can be defined by the similarity in the following way:

Count: For any object x in a possible world w and any object y in a possible world v , x is the counterpart of y in w relative to one respect iff x in w is similar to y in v than any other object in w relative to this respect.

The fictional person-idea is the counterpart of p in $@$ since it is the object most similar to p in $@$ relative to the respect of fiction. Thus, statement 3) is true if it is uttered in a fictional world but evaluated in the actual world.

Consider ordered pairs of worlds $\langle v, w \rangle$ where w can either be the same or not the same as v . The truth condition of metafictional statements like 3) can be given in the following way:

True: A sentence s is true to $\langle v, w \rangle$ iff s is true if it is uttered in v and evaluated in w .

Met: A metafictional statement m is true iff m is true to every $\langle v, @ \rangle$ where v is a fictional world and $@$ is the actual world.

The above interpretation of metafictional statements avoids the “wrong kind of object” problem while holding the core thesis of possibilism. What is picked out by fictional names are still concrete objects in non-actual possible worlds, but it is a fictional-person idea that is ascribed to metafictional properties. The fictional-person idea of Sherlock Holmes is just the right kind of object to be the subject of “being a fictional character created by Conan Doyle.”

The interpretation also allows “Sherlock Holmes” in a metafictional context to have the same meaning as “Sherlock Holmes” in a parafictional context. The truth conditions for parafictional statements can be adjusted within a two-dimensional framework to align with metafictional ones.

Para 2: A parafictional statement n is true iff n is true to every $\langle v, v \rangle$ where v is any fictional world.

The above condition says a parafictional statement is true iff it is true when uttered and evaluated in any fictional world. If statement 2) is uttered in a fictional world, “Sherlock Holmes” picks out a real person p , and there would be a complete proposition like $\langle p, \text{a private detective} \rangle$ expressed. Obviously, the proposition is true when evaluated in the fictional world. “Sherlock Holmes” works in the same as it does in the metafictional context. The

meaning of "Sherlock Holmes" can be seen as a partial function from possible worlds to individuals in each possible world.

3. Fictional Names and Counterpart Relations

In this section, I want to address one potential worry about my solution. The worry is why a fictional-person idea is the counterpart in the actual world instead of other concrete objects. If, in an utterance world, Darwin is just like how Holmes is depicted in the stories, including having the name "Sherlock Holmes," then "Sherlock Holmes" would denote Darwin in this utterance world. In this case, why the counterpart of Darwin in the actual world is not Darwin?

The reason is, when evaluating metafictional statements, what matters is the objects' relation with fictional names. Darwin might be related to "Sherlock Holmes" in some fictional worlds, for he could be named "Sherlock Holmes," but Darwin is definitely not related to "Sherlock Holmes" in the actual world. The naming baptism of "Charles Darwin" is irrelevant to "Sherlock Holmes." In contrast, the fictional-person idea of Sherlock Holmes emerges from a fictitious use of "Sherlock Holmes." Conan Doyle told a story as if "Sherlock Holmes" denotes someone. This is not a genuine naming baptism, but it shares something in common with the genuine naming baptism of "Sherlock Holmes" in fictional worlds. The practice of naming related to the objects constitutes a counterpart relation between the possible Darwin and the fictional-person idea in the actual. Thus, even in the above case, the counterpart in the actual world should be a fictional-person idea.

References

Chalmers, David J. (2006) "The Foundations of Two-Dimensional Semantics", in: Manuel Garcia-Carpintero & Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press. 55-140.

Everett, Anthony & Schroeder, Timothy (2015) "Ideas for stories", in: Stuart Brock & Anthony Everett (eds.), *Fictional Objects*. Oxford University Press.

Everett, Anthony (2013) *The Nonexistent*. Oxford University Press.

Klauk, T. (2014) "Zalta on Encoding Fictional Properties", *Journal of Literary Theory*, 8(2), 234-256.

Kripke, S. A. (2013) *Reference and Existence: the John Locke Lectures*. Oxford University Press, USA.

Kroon, Fred and Alberto Voltolini, "Fictional Entities." *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <<https://plato.stanford.edu/archives/fall2023/entries/fictional-entities/>>.

Lewis, D. K. (1968) "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy* 65 (5), 113-126.

Lewis, D. (1978) "Truth in Fiction", *American Philosophical Quarterly* 15 (1), 37-46.

Murday, Brendan (2011) "Two-Dimensionalism and Fictional Names", in: Franck Lihoreau (ed.), *Truth in Fiction*. Ontos Verlag. 43-76.

Recanati, F. (2018) "II—Fictional, Metafictional, Parafictional", *Proceedings of the Aristotelian Society*, Vol. 118, No. 1, 25-54. Oxford University Press.

Semeijn, M., & Zalta, E. N. (2021) "Revisiting the 'Wrong Kind of Object' Problem", *Organon F*, 28 (1), 168-197.

Stalnaker, R. C. (1978) "Assertion", in: *Pragmatics*. Brill, 315-332.

Thomasson, A. L. (1999) *Fiction and Metaphysics*. Cambridge University Press.

Voltolini, A. (2019) "Varieties of Fiction Operators", *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications*, 199-210.

Walton, K. L. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.

A Heideggerian Reading of the *Tractatus*

David Lindeman (Washington D.C., USA)

Abstract

According to the orthodox reading, the nonsensical but elucidatory sentences of the *Tractatus*—elucidations—intend but fail to say what sentences with sense and the senseless propositions of logic show. According to the resolute reading, all nonsense is on a par. I provide a novel alternative reading. As with the orthodox reading, a distinction is drawn between mere nonsense and elucidatory nonsense. Against the orthodox reading, it is maintained that Wittgenstein's elucidations show, reveal, or disclose aspects of reality of which we cannot speak. This idea is elaborated with reference to Heidegger's conception of truth as ἀλήθεια. A distinguishing mark of this interpretation is that it takes seriously the mysticism of the *Tractatus*.

I

The world is the totality of facts. So we read at the start of the *Tractatus*. It is an echo of the opening line of Schopenhauer's great work: *The world is my representation*—the totality of my thoughts. At first blush, this might look opposed to Wittgenstein's talk of the totality of thoughts *picturing* the world. But his states of affairs are thinkables: positions in logical space, the totality of which is reality, the sum of what is logically possible—the sum of what can be thought (TLP 1.13, 2.04, 2.06, 3.001, 3.01, 3.02, 3.03, 3.4, etc.). The state of affairs thought out in thinking the sense of a propositional sign, if obtaining, is a fact. If it is a fact, the thought is true. But the fact is not something independently specifiable, that is, specifiable outside the thought the sense of which is that state of affairs. The totality of true thoughts is thus a picture of the world *as pictured*.

This interpretation of the *Tractatus* stands opposed to a “realist” interpretation, according to which analysis of language reveals the ontological structure of an independently constituted reality. But it is equally opposed to an “anti-metaphysical” interpretation, according to which analysis is a purely language-internal logical investigation. There is a middle way, neglected in the debate just alluded to, and this is that the *Tractatus* is a treatise in the phenomenological tradition: it is at once a logical and a metaphysical investigation.

Phenomenology, as Heidegger argues, *is* metaphysics; and the method of this metaphysics for the early Wittgenstein is analysis. It is an analytic

phenomenology. This is not an uncritical metaphysics. Being close kin to Kant's critique of reason, Wittgenstein's philosophy is (as he says) a "critique of language" (TLP 4.0031). The task is to reveal the nature of the proposition, that is, as already indicated, the nature of thought, which is constitutively logical, and so the essence of all representation (TLP 5.61; 3.03, 3.031, 5.4731)—the essence of the world (TLP 5.4711).

II

Gadamer's famous dictum is called to mind: *Being that can be understood is language*. The remark occurs in part III of *Truth and Method* in a passage opening with the observation that in language "I and world [...] manifest their original belonging together" (469). Similar formulations are found throughout the work of Heidegger, who in his "Letter on Humanism" poetically identifies language as the "house of Being". In *Being and Time*, the original belonging is expressed as identity: "Dasein is its world existingly" (BT 416). One cannot fail to be struck by how close that formulation comes to one found in those passages of the *Tractatus* in which Wittgenstein addresses the question "how much truth there is in solipsism": "I am my world." (TLP 5.63)

But how significant are these surface similarities? "Dasein", Heidegger's replacement for "the self" or "I", "an entity which does not just occur among other entities", he tells us, but is rather "ontically distinguished by the fact that, in its very Being, that Being is an *issue* for it", is one always already in the world (BT 32). This appears to be contradicted by remarks made by Wittgenstein following the one just quoted:

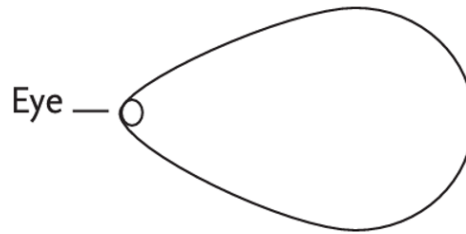
5.632 The subject does not belong to the world: rather, it is a limit of the world.

5.633 Where *in* the world is a metaphysical subject to be found?

You will say that this is exactly like the case of the eye and the visual field. But really you do *not* see the eye.

And nothing *in the visual field* allows you to infer that it is seen by an eye.

5.6311 For the form of the visual field is surely not like this



The metaphysical subject is not in the world. Surely, then, it is not Dasein, which is in the world? But by this I take Wittgenstein to mean that the self is not an *object* in the world—as objects are objects in the world, that is just to say it is not an object. In remarks in the *Notebooks* found among those that found their way into this stretch of the *Tractatus*, Wittgenstein remarks simply: “The I is not an object.” (NB 7.8.16) For all that, it is. The I is not, as Heidegger agrees, and now putting it his way, another entity present-at-hand. Like the eye to the visual field, it is a presupposition, a precondition—in Kantian terminology, a transcendental condition for the possibility—of anything’s showing up (being present) in the world. What brings the self into philosophy is the fact that “the world is my world” (TLP 5.641).

III

What world? Not the phenomenal world, if this is to be contrasted with a noumenal world. Like the German idealists before them, neither Wittgenstein nor Heidegger draws the Kantian distinction. Phenomenology studies the phenomenon in a more original sense, as Heidegger emphasizes, one going back to the Greek *φαινόμενον*: viz. that which appears, manifests, shows itself, comes to light (see BT ¶7): “the *φαινόμενα* or ‘phenomena’ are the totality of what lies in the light of day or can be brought to the light”. This bringing to light, what is in some sense always already there before us, is accomplished by phenomenology.

Of course, “phenomenology”, as Heidegger notes, has two components: “phenomenon” and “logos”. For Wittgenstein, the logos of the phenomenon, the method of his phenomenology, what brings to light the essence of the phenomenon (what brings being into the intelligibility of Dasein, as Heidegger would put it), is logic. Logical analysis is the critique of language.

What language? At a first pass, *my* language, the limits of which are the limits of *my* world (TLP 5.6). That the world is my world “is manifest in the fact that

the limits of *language* (of that language which alone I understand) mean the limits of *my world*” (TLP 5.62). We might put it the other way round: My world is *the world*. It is not mine alone. Likewise, by “that language which alone I understand” is meant not the language which only I understand but the only language I understand, the language of thought—the only language there is to be understood, the logic of which excludes no possibilities. “Logic pervades the world: the limits of the world are also its limits.” (TLP 5.61) This, he says, is “the key to the problem, how much truth there is in solipsism” (TLP 5.62).

5.64 Here it can be seen that solipsism, when its implications are followed out strictly, coincides with pure realism. The self of solipsism shrinks to a point without extension, and there remains the reality co-ordinated with it.

The reality—or world—, we might say, *co-constituted* with it. *Being-in-the-world* is fundamentally a *unity*: “World-understanding is Dasein-understanding is self-understanding”, Heidegger writes in *The Basic Problems of Phenomenology* : “Self and world are the basic determinations of the Dasein itself in the unity of the structure of being-in-the-world.” (297) What appeared to be a difference between Wittgenstein and Heidegger, on closer analysis, is a fundamental agreement.

IV

The bringing to light of the phenomenon is again the function of logos: “The λόγος lets something be seen (*φαίνεσθαι*), namely, what the discourse is about” (BT 56).

The ‘Being-true’ of the λόγος as ἀλήθεια means that [...] the entities of *which* one is talking must be taken out of their hiddenness; one must let them be seen as something unhidden (ἀλήθής); that is, they must be *discovered*. (BT 56-7)

On Wittgenstein’s view, it is again analysis that achieves this discovering, uncovering, unconcealing, disclosing. “There is no ontology *alongside* a phenomenology”, Heidegger writes in *The History of the Concept of Time: Prolegomena*. “Rather, *scientific ontology is nothing but phenomenology*” (75).

And now the promise of taking logic as first philosophy, as Wittgenstein does, is precisely that ontology is made scientific.

V

We have said the critique of language is *scientific*. It is not, however, according to Wittgenstein, a natural science.

4.111 Philosophy is not one of the natural sciences.

(The word ‘philosophy’ must mean something whose place is above or below the natural sciences, not beside them.)

On Wittgenstein’s conception, the natural sciences are concerned with objects in the world. “Logic”, by contrast, he unequivocally states, “is transcendental” (TLP 6.13). It deals with the forms immanent in our comprehension of the states of affairs comprising the objects of the world. (A concern with Being, or the meaning of Being, likewise, is not concern for an object—not a concern with beings, entities, but the possibility of their showing up for us.) In this way its concern is with the conditions for the possibility of natural science.

4.11 The totality of true propositions is the whole of natural science (or the whole corpus of the natural sciences).

It follows that there are no philosophical propositions, that is, propositions of philosophy, if propositions are propositions with sense, and propositions with sense concern states of affairs, the obtaining states of affairs comprising the objects of the world. “A philosophical work’, instead, ‘consists essentially of elucidations” (TLP 4.112).

VI

Part of what the elucidations of philosophy achieve is a setting of limits.

4.113 Philosophy sets limits to the much disputed sphere of natural science.

Again, the scope of natural science is restricted to the objects of the world. Its facts, *in toto*, are the world. What it does not represent are the conditions for the possibility of this representation.

4.12 Propositions can represent the whole of reality, but they cannot represent what they must have in common with reality in order to be able to represent it—logical form.

Logic is again transcendental.

5.552 The ‘experience’ that we need in order to understand logic is not that something or other is the state of things, but that something *is*: that, however, is *not* an experience.

Logic is *prior* to every experience—that something is *so*.

It is prior to the question ‘How?’, not prior to the question ‘What?’

5.5521 And if this were not so, how could we apply logic? We might put it in this way: if there would be a logic even if there were no world, how then could there be a logic given that there is a world?

VII

That there is something, that there is anything, that the world exists – this is not explained.

6.371 The whole modern conception of the world is founded on the illusion that the so-called laws of nature are the explanations of natural phenomena.

6.372 Thus people today stop at the laws of nature, treating them as something inviolable, just as God and Fate were treated in past ages. And in fact both are right and both wrong: though the view of the ancients is clearer in so far as they have a clear and acknowledged terminus, while the modern system tries to make it look as if *everything* were explained.

The metaphysical self, too, as it is not an object, is not an object of scientific inquiry, and therefore not part of what is explained by the natural sciences. (This is supposed to be revealed by logical analysis; see the 5.54s.) The natural sciences, in brief, do not concern “what is higher” (TLP 6.42, 6.432).

VIII

What is higher? “What do I know about God and the purpose of life?” Wittgenstein asks himself in the *Notebooks*.

I know that this world exists.

That I am placed in it like my eye in its visual field.

That something about it is problematic, which we call its meaning.

That this meaning does not lie in it but outside it.

The second line receives more exact formulation in the *Tractatus*, where the last line also recurs in altered form:

6.41 The sense of the world must lie outside the world. In the world everything is as it is, and everything happens as it does happen: *in* it no value exists—and if it did exist, it would have no value.

Likewise in 6.4312: “The solution of the riddle of life in space and time lies *outside* space and time. (It is certainly not the solution of any problems of natural science that is required.)” *Why?*

6.432 *How* things are in the world is a matter of complete indifference for what is higher.

God does not reveal himself *in* the world.

6.4321 The facts all contribute only to setting the problem, not to its solution.

IX

“It is absurd to say ‘Science has proved that there are no miracles.’” Wittgenstein remarks in his “Lecture on Ethics”. “The truth is that the scientific *way of looking at a fact* is not the way to look at it as a miracle.” (my emphasis) And is looking at it as a miracle “seeing it aright”? (TLP 6.54) This is presumably an alteration of an individual’s over-all attitude to the one world, Heidegger’s web of significances, Gadamer’s horizon of meaning, the world the limits of which for Wittgenstein are those of logic. It is an attitude to the world as a whole—to the world taken as a limited whole: “*Feeling* the world as a limited whole—it is this that is mystical” (TLP 6.45, emphasis added). Who will deny that the mood of the *Tractatus* is a mystical one?

Mood, what “arises out of Being-in-the-world, as a way of such being” (BT 176): *Stimmung*, as in the tuning of an instrument—being differently tuned one is attuned to, resonates with, different things; different things show up for one. “The world of the happy man is different from that of the unhappy man.” (TLP 6.43) The world of the mystic a different world from that of the scientist qua scientist. And that is the dominant attitude of the present day. “People nowadays think that scientists exist to instruct them, poets, musicians, etc. to give them pleasure.” Wittgenstein remarks in *Culture and Value*. “The idea that these have something to teach them – that does not occur to them.”

Perhaps in fact Wittgenstein’s happy man *is* the mystic. Happiness comes up in the *Notebooks* in connection with aesthetic appreciation and this with the miracle of existence, which inspires mysticism.

Aesthetically, the miracle is that the world exists. That what exists does exist.

Is it the essence of the artistic way of looking at things, that it looks at the world with a happy eye? (NB 20.10.16)

For there is certainly something in the conception that the end of art is the beautiful.

And the beautiful *is* what makes happy.

We read in the *Tractatus* that aesthetics and ethics are one and the same (TLP 6.421). We remember also that Wittgenstein remarked once that the point of

the book was ethical, that the most important part of the book is the part he did not write—the silences.

X

“To pray is to think about the meaning of life”, Wittgenstein writes in the *Notebooks*—that is, provided the identification made on the same day, an identification which appears also in the *Tractatus*: to think of the meaning of the world (NB 11.6.16, TLP 5.621). I take this to be the same as thinking of the meaning of Being: to pray, then, is to ask Heidegger’s *Seinsfrage*—the question of the meaning of Being. It is sometimes said that “Being” for Heidegger was a substitute for “God”. Here Wittgenstein performs the substitution the other way: “The meaning of life, i.e. the meaning of the world, we can call God.”

But again, God does not reveal himself *in* the world. So Wittgenstein will say that there is really no question here, that is, no question with sense, and likewise no answer with sense.

- 6.5 When the answer cannot be put into words, neither can the question be put into words.
 The *riddle* does not exist.
 If a question can be framed at all, it is also *possible* to answer it.

For “propositions can express nothing that is higher” (TLP 6.42)—that is, nothing of anything outside the world. Thus:

- 6.521 The solution of the problem of life is seen in the vanishing of the problem.

Whereas Kant sets out in *The Critique of Pure Reason* to deny knowledge to make room for faith, Wittgenstein sets out in the *Tractatus* to deny thought, to delimit the bounds of sense, to make room for silence.

XI

The later Heidegger, as Rorty observes, “wanted to discover a language that was as close to silence as possible”. Wittgenstein once remarked that he liked the idea of a silent religion. The silence where there was once a question. Is that the answer?

6.522 There are, indeed, things that cannot be put into words.
They *make themselves manifest*. They are what is mystical.

We provide a setting of limits, by working outward from what can be said to what cannot be said; a clearing, a lighting, for the witness of being.

XII

The picture we arrive at is this. Talking is producing pictures, images. Images are of objects. Truth, being, logical form, etc.—what is higher, in brief, these are not objects. So we can't make pictures of them. We can't talk about them. Problem: We just spoke of these things, and purported to say true things about them – for example, that they are not objects. So we need some other sense in which we can talk about them, some sense in which we can say true things about them. Solution: We can recognize that elucidations are true in the sense Heidegger recognizes; that they are sayings in the sense of disclosures: they make manifest, bring to light, show the things of which, in Wittgenstein's proprietary sense, we cannot speak. There *are* truths *in this sense* (truth in the sense of ἀλήθεια, a more original sense of truth) about “what is higher”—what we might call higher truths. It is neither the orthodox nor resolute reading but a Heideggerian reading, what might equally be called the esoteric reading.

References

Gadamer, H. (2006) *Truth and Method*. J. Weinsheimer and D.G. Marshall (trans.), New York: Continuum.

Heidegger, M. (1982) *The Basic Problems of Phenomenology*. A. Hofstadter (trans.), Bloomington: Indiana University Press.

Heidegger, M. (1985) *The History of the Concept of Time: Prolegomena*. T. Kisiel (trans.), Bloomington: Indiana University Press.

Heidegger, M. (1993) "Letter on Humanism", F. A Capuzzi and J. Glenn Gray (trans.). in: D. F. Krell (ed.) *Martin Heidegger: Basic Writings*, Revised and Expanded Edition, London: Routledge, 1993, 217–65.

Heidegger, M. (2001) *Being and Time*. J. Macquarrie and E. Robinson (trans.), Oxford: Blackwell.

Rorty, R. (1991) "Pragmatism Without Method", in: *Objectivity, Relativism, and Truth: Philosophical Papers*. Cambridge: Cambridge University Press.

Wittgenstein, L. (1961) *Notebooks 1914–1916*. G. H. von Wright and G. E. M. Anscombe (eds.), Oxford: Blackwell.

Wittgenstein, L. (1965) A Lecture on Ethics. *The Philosophical Review*, 74: 3–12.

Wittgenstein, L. (1974) *Tractatus Logico-Philosophicus*. D.F. Pears and B.F. McGuinness (trans.), New York: Routledge.

Wittgenstein, L. (1980) *Culture and Value*. G.H. von Wright (ed.), P. Winch (trans.), Oxford: Blackwell.

A Private Language Argument against Phenomenal Concepts

Yaojun Lu (Syracuse, USA)

Abstract

In the popular account of our epistemic relation with phenomenal consciousness, we know what we experience by categorizing it with phenomenal concepts, a type of concepts that we acquire by introspection. Many philosophers of mind claim that by introspecting a phenomenal experience we thereby produce a copy of it and use that as a concept. I argue that there are no phenomenal concepts and this account conflates a concept with its referent. I appeal to Wittgenstein's Private Language Argument, that for a concept to bear significance and thus for the knowing to be a coherent notion, there must be rules and criterion of correction regulating the use of the concept. Since phenomenal concepts are acquired purely by introspection, there is no such rule or criterion of correctness. I will use Frank Jackson's thought experiment of Mary as the discussion material and respond to David Papineau's recent defense of phenomenal concepts.

1. Phenomenal Concepts

Phenomenal concepts are a recent invention as part of the so-called "phenomenal concept strategy" against anti-physicalist arguments especially Jackson's (1982, 1986) Knowledge Argument (KA) (see review by Stoljar, 2005). KA appeals to the controversial principle that an epistemic gap must arise from an ontological gap, and it claims to have found an explanatory gap in the thought experiment of Mary. Mary is allegedly omniscient *de dicto* about possibly all physical facts related to color experience but has never been exposed to colored objects. She grows up and stays within a black-and-white room, and she achieves her *de dicto* omniscience from descriptions of all the physical facts. Jackson's intuition is that when Mary leaves her room and sees a ripe tomato, she would be surprised and learn a new piece of knowledge, of what it is like to see red. Jackson takes Mary's cognitive change to indicate an epistemic gap, as her omniscience is unable to a priori entail her new knowledge. Jackson concludes that this epistemic gap between knowledge of phenomenal experiences and knowledge of physical facts arises from the ontological gap between phenomenal experiences and physical facts.

The phenomenal concept strategy is a moderate response to KA. Whereas some hardcore physicalists (e.g., Churchland, 1985; Dennett, 1991) deny Jackson's intuition thus deny the epistemic gap, the so-called "non-reductive" physicalists acknowledge the epistemic gap but reject the principle linking the epistemic and the ontological. They propose an alternative diagnosis of the

epistemic gap, that it arises from a conceptual gap. Just as in the Kripkean cases such as Phosphorus and Hesperus, they argue that the gap is on the level of concepts not on the level of reference. They thus argue that what Mary lacks in the room and what she acquires upon seeing red is a new concept that refers to a physical property that she previously referred to via a physical concept. Thus, when she has the first red experience, she acquires a new concept RED_q which co-refers with her old concept RED_p. There is no a priori link between the two, as the cognitive significance of RED_q is the phenomenal character whereas the cognitive significance of RED_p is all the relevant physical information.

Interestingly, while physicalists and anti-physicalists disagree on whether the phenomenal concept strategy is an adequate response, they agree that we possess a special type of concepts the acquisition of which requires having the right experience. For example, Papineau claims that “phenomenal concepts are compound terms, formed by entering some state of perceptual classification or re-creation into the frame provided by a general experience operator ‘the experience: ---’. [...] Very roughly speaking, we refer to a certain experience by producing an example of it.” (2002, p. 116) Chalmers claims that “The clearest cases of direct phenomenal concepts arise when a subject attends to the quality of an experience and forms a concept wholly based on the attention to the quality, “taking up” the quality into the concept.” (Chalmers, 2003/2010, p. 267) They agree that Mary’s cognitive change occurs when she introspects and acquires a new concept.

There have been some criticisms of phenomenal concepts. One appeals to semantic externalism (Ball, 2009; Tye, 2009). Accordingly, just as whether a person possesses a certain concept depends on external factors, pre-release Mary possesses the concept of red experience just as color-blind or fully blind people do not. What Mary and these people lack is the experience and things that depend on it, not the concept. Another criticism focuses on the requirement that concepts must be storable mental representations but phenomenal concepts are at best demonstratives which cannot be stored (Prinz, 2007). Perhaps the most iconic criticism is that concept and its referent are confused in the introspective acquisition of phenomenal concepts. Leibniz complained that Locke’s notion “idea” is ambiguous between a sensation and the power that causes a sensation (Bennett, 1996). Hume famously said that “If

you tell me, that any person is in love, I easily understand your meaning, and form a just conception of his situation; but never can mistake that conception for the real disorders and agitations of the passion.” (2007, §II)

I want to show that Wittgenstein’s Private Language Argument is another, if not the strongest, criticism of phenomenal concepts. It entails that there cannot be any phenomenal concept because it would have no significance. Now I turn to a brief exegesis of this argument.

2. The Private Language Argument

The Private Language Argument (PLA) is part of Wittgenstein’s criticism of the naïve view of rule following, and rule following is the central for his ideas of language games and meaning as use. According to the naïve view, following a rule consists in consulting a consciously registered sentence or a picture representing the rule. At many places, Wittgenstein made the point that such a conscious state would require interpretation if to offer guide (PI 2009: §22, §31, §34, §60, §85, §139). The naïve response is that the right interpretation of the sentence or picture may at the same time occur to consciousness also in a form of a sentence or a picture (PI 2009: §139, §141, §151). Before eventually presenting PLA, Wittgenstein gave two distinct answers. One is that there is no such conscious state. He used the case of understanding, i.e., realizing a rule, and reading to show that following a rule is a disposition and our language would suffer category mistake if we treat it as an event (PI 2009: §149, §151, §154, §157) (c.f., Ryle, 1949/2009). The other is that a paradox would follow. After repeating the naïve view (PI 2009: §169, §173, §178, §183, §191) and discussing the example of a child who applies the operation of +2 only to numbers smaller than 1,000 (PI 2009: §185), Wittgenstein argued that if the correct use of a rule must be mediated by a conscious interpretation, then paradoxically every action is compatible with any rule and thus no rule can be used incorrectly (c.f., Kripke, 1982). He concluded that following a rule is a practice such that whether one follows a rule is not decided by the person thinking that they are following it (PI 2009: §202).

PLA is presented through several famous cases, where concepts are introduced privately as names of sensations which entails that the rules governing their use are also private. One is a genius child who “invents a name for [toothache]

by himself” (PI 2009: §257) and one is the diarist who names a certain sensation “S” by concentrating on the sensation (PI 2009: §258). Wittgenstein’s presentation spans many sections, but he immediately commented that such private concepts “have no criterion of correctness. One would like to say: whatever is going to seem correct [...] is correct. And that only means that here we can’t talk about ‘correct’.” (PI 2009: §258) The naïve response is that we can use memory as the “independent authority” to ground correctness, but Wittgenstein pointed out the problem is that judging whether the memory is correct lacks its own criterion of correctness (PI 2009: §265). Since there is no correctness in the use of such concepts, it makes no sense to say they bear any significance. Such concepts would not mean anything, just as a knob or a wheel that has no effect in a machine (PI 2009: §270, §271).

There are at least two common misunderstandings of a private language. One is that it is any language that refers to sensations. Under this understanding, PLA is contradicted by our own case because we have a language that refers to private sensations and Wittgenstein clarified that this is not what he meant (PI 2009: §243). A private language consists of concepts that refer to things that can be known by the user, and sensations are not private in this sense, according to Wittgenstein. He acknowledged that our sensations are private but challenged the notion that sensations are private in terms of our private knowing of them. For him, everyone can know that I am in pain, and my pain is private only in the sense that I am the only one that can have them, which is a truism (PI 2009: §246, §248). The other misunderstanding is that a private language is any language that can have only one user (c.f., Ayer, 1954). This would make private languages obviously possible and often actual. There have been, possibly are, and will be languages that lose their users to a point that there is left only one user. Also, someone can encrypt a language to be intelligible to only one user. Such languages do not feature what philosophers call “logical privacy” as explained above.

Wittgenstein used the famous beetle in the box case to characterize a private language. If everyone has a box that no one else can look inside and everyone has a concept “beetle” to refer to what is inside their own box, this makes a private concept. Such concept bears no significance, because everyone’s “beetle” might be different, each “beetle” might be a changing object, or there might be nothing at all (PI 2009: §293). Wittgenstein related this back to her

earlier expressivist thesis, that the relation between a sensation word and the sensation is that the word replaces the natural expressions of the sensation (PI 2009: §244). The relation is not that we attend to the sensation and give it a name, because if so then the sensation is not connected to our language game just like the beetle is (PI 2009: §293). It becomes an ornament knob or a freely spinning wheel that he spoke of earlier. Here Wittgenstein was focusing on the establishment of the reference relation, as he clarified that he was not a behaviorist (PI 2009: §307).

The so-called “stage-setting” interpretation, it seems to me, concisely captures PLA especially the beetle in the box case. In the very section that Wittgenstein imagined the genius child, he said “But what does it mean to say that he has ‘named his pain’? How has he managed this naming of pain? [...] When one says ‘He gave a name to his sensation’, one forgets that much must be prepared in the language for mere naming to make sense.” (PI 2009: §257) Imagine that I simply say “I name it ‘beetle’”—there is no box that I raise because it is something completely private. It is as if the sensation has no manifestation whatsoever, so no one else can know what I mean (PI 2009: §256). Moreover, not even I can know what it is. There is no information available to me such as “I experience it when I focus my eyes on a ripe tomato” or “it gives the feeling of warmth”. The “logical privacy” makes it nonsense that there is something named, so it is not an issue of correct memory (c.f., diagnosis by Wrisley, 2011).

3. The Logical Privacy of Phenomenal Concepts

The logical privacy of phenomenal concepts can be revealed in two perspectives. One is that Mary’s newly acquired concept RED_q has no a priori link to any physical information that she already knows. If there was, then it would be possible for her to possess concept RED_q while in the room hence would not learn anything new upon release. It is after her acquiring a phenomenal concept that she links it to relevant physical information, so the cognitive significance of her phenomenal concept, if it has any, is the phenomenal character of an experience and nothing else. Another perspective is that Mary’s cognitive change occurs in her introspection. Before having a red experience, her omniscience of physical information tells her that she is about to experience what people call “red” which is produced by light of

certain wavelengths and typically instantiated by certain objects. Other than introspection, there is nothing else that she can learn to build the association between the experience and her omniscience. Physicalists would say that she builds a posteriori identity, whereas anti-physicalists say that she builds correlation between two properties.

Nida-Rümelin's (1995) variation of Jackson's thought experiment, the one that involves a character Marianna, clearly shows the logical privacy of phenomenal concepts. Marianna is in every aspect the same as Mary, except that before she sees any colored object she is first exposed to colored patches. According to the advocates of phenomenal concepts, she would have acquired the phenomenal concept which Mary acquires upon leaving her room. Before leaving her room, she wonders what is "red" that everyone talks about. Marianna obviously does not acquire a demonstrative "this", because she has already had that. She allegedly acquires a new concept—maybe we can call Q1—that enables her to think about color experience independently of a demonstrative. When Marianna sees colored objects, her omniscience enables her to match the object and the experience, then she would know that "Q1 is what people call 'red'."

Equivalently, we may imagine that Marianna acquires phenomenal concepts by being stimulated via brain-computer interface. This makes it easier for us to apprehend Marianna's epistemic situation. Imagine that right now you suddenly undergo an experience which is nothing like anything that you have known or experienced, so not only that you cannot explain it to others but more importantly you yourself do not know what it is. It is not even dependent on your control of any sense organ, so you do not even know which modality it belongs to.

Now we can go back to think again of how Marianna can give the name Q1. It may seem that her new experience is the only bizarre thing that she tries to name, so the naming is established by the uniqueness. But the problem is that there is no rule governing her future use of this name. It would be meaningless to say that her subsequent use of it is correct or not. Obviously, there is not even a correct memory to provide criterion. We can see this problem more clearly by turning to the growth of Marianna's conceptual repertoire. Suppose she then sees a green patch and names it Q2. Of course she does not know that

it is green as we mean it, and worse still, she does not and cannot know if Q1 is Q1 and Q2 is Q2. At best she may know that they are different, but even this is questionable. Therefore, Q1 and Q2 are not names of anything. It is not that Marianna forgets what they name; they are names of indistinguishable things to begin with. There are indeed two different things to be named, but Q1 and Q2 are just like word “beetle” in Wittgenstein’s story. They are two names that have no meaning.

4. Papineau’s Defense of Phenomenal Concepts

David Papineau has recently (2011) shown that phenomenal concepts are indeed incompatible with PLA but he argues that the fault is in PLA. He recognizes that one way to possess a concept is to subscribe to certain rules, as Wittgenstein argued. But he then argues that there is an alternative way to possess a concept which “does not involve rules of any kind”. The idea is that as a brute fact we may possess some natural, ready-made concepts that “are designed to lock onto entities in our environment”. He claims that no rule is involved, as the user of such concepts may be “hopeless” at using them correctly (p. 182). His argument involves some misinterpretations of PLA such as that it is about communication, but I shall focus on his rule-less account of phenomenal concepts.

The obvious reply is that Papineau has conflated there being rules with the user knowing the rules. The core of PLA is the conceptual link between there being rules and there being criterion of correct use, and Papineau’s possibility is the one where the user has no knowledge of the rules but not one where there are no rules. If there really are no rules, then there would be no criterion of correctness, and Papineau’s claim of “hopeless” users would be meaningless. In real life, many native users of a language cannot explain the grammar of their language and many are “hopeless” at using the language, but we do not conclude that this is a grammarless language. We can further look at perhaps some ready-made concepts. The most basic concepts such as “time” and “the world” seem to be the best candidates for Papineau’s rule-less concepts, and possibly some people are hopeless in using them. But we do not say that no rules govern the use of such concepts.

We should then ask how Papineau's ready-made concepts are hooked onto their referents. The most plausible account seems to be evolutionary, that such concepts are fit through the process of natural selection. He also quote Millikan (2000) who holds a biological account of concepts to make sense of his idea. I do not see how in the evolutionary account there are no rules governing the use of the ready-made concepts. Furthermore, the evolutionary account is not an alternative to Wittgenstein's rule-based account; if anything, it supports Wittgenstein's account. Wittgenstein believed that sensation words are expressive and the evolutionary story of that would be that replacing natural expressions with words provides fitness.

5. Conclusion

A clear examination of PLA and phenomenal concepts shows a grave misconception of our epistemic relation with phenomenal consciousness. Many philosophers believe that we are literally thinking about a phenomenal experience when we recall it in consciousness. They call this episode of recalled experience "thinking", because they believe that merely attending to a phenomenal experience can create a concept that predicates this very experience. They thus conflate a mental undergoing and its categorization, and in general the receptive mind and the intellect. PLA shows that phenomenal concepts conceived as such are not concepts at all, as they do not bear any significance. We do have concepts of phenomenal experiences, but they are not the kind whose acquisition or even constitution requires having the right experience.

References

- Ayer, Alfred J (1954) "Can There Be a Private Language?", *Proceedings of the Aristotelian Society, Supplementary Volumes*, 28, 63–94.
- Ball, Derek (2009) "There Are No Phenomenal Concepts", *Mind*, 118(472), 935–962.
- Bennett, Jonathan (1996) "Ideas and Qualities in Locke's "Essay."", *History of Philosophy Quarterly*, 13(1), 73–88.
- Chalmers, David J. (2010) *The Character of Consciousness*. Oxford University Press.
- Churchland, Paul M. (1985) "Reduction, Qualia, and the Direct Introspection of Brain States", *The Journal of Philosophy*, 82(1), 8–28.

- Dennett, Daniel C. (1991) "'Epiphenomenal' qualia?", in: Y. Nagasawa, P. Ludlow, & D. Stoljar (Eds.), *There's something about Mary: Essays on phenomenal consciousness and Frank Jackson's knowledge argument*, 127–136. Little, Brown.
- Hume, David (2007) *An Enquiry concerning Human Understanding* (P. Millican, Ed.). Oxford University Press.
- Jackson, Frank (1982) "Epiphenomenal qualia", *Philosophical Quarterly*, 32(April), 127–136.
- Jackson, Frank (1986) "What Mary Didn't Know", *Journal of Philosophy*, 83(May), 291–295.
- Kripke, Saul A. (1982) *Wittgenstein on Rules and Private Language* (Vol. 2, Issue 2, p. 342). Harvard University Press.
- Millikan, Ruth G. (2000) *On Clear and Confused Ideas: An Essay about Substance Concepts* (Illustrated edition). Cambridge University Press.
- Nida-Rümelin, Martine (1995) "What Mary Couldn't Know: Belief About Phenomenal States", in: T. Metzinger (Ed.), *Conscious Experience* (pp. 219–241). Imprint Academic.
- Papineau, David (2002) *Thinking about Consciousness*. Clarendon Press.
- Papineau, David (2011) "Phenomenal Concepts and the Private Language Argument", *American Philosophical Quarterly*, 48(2), 175.
- Prinz, Jesse (2007) "Mental Pointing: Phenomenal Knowledge Without Concepts", *Journal of Consciousness Studies*, 14(9–10), 184–211.
- Ryle, Gilbert (2009) *The Concept of Mind: 60th Anniversary Edition*. Routledge.
- Stoljar, Daniel (2005) "Physicalism and Phenomenal Concepts", *Mind and Language*, 20(2), 296–302.
- Tye, Michael (2009) *Consciousness Revisited: Materialism Without Phenomenal Concepts*. MIT Press.
- Wittgenstein, L. (2009) *Philosophical Investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.; Rev. 4th ed). Wiley-Blackwell.
- Wrisley, George (2011) "Wherefore the Failure of Private Ostension?", *Australasian Journal of Philosophy*, 89(3), 483–498.

Non-representationalism and Anti-Metaphysical Arguments

Shiman Luo (Vienna, Austria)

Abstract

Metaphysical placement problems are problems with “placing” moral facts (or properties), mathematical facts, or modal facts, and alike, in the world as we know it. Such problems can be provoked by truthmaking questions, i.e., questions of what, if any, makes the sentences and thoughts about a given area of facts true. Non-representationalists reject truthmaking questions to avoid the relevant way of yielding metaphysical placement problems. In this paper, I represent and examine three anti-metaphysical arguments in the literature, provided by paradigmatic non-representationalists including Simon Blackburn and Huw Price. I discern two different motivations behind these anti-metaphysical arguments. Then, I argue that the second motivation is more justified than the first. The aim is to elucidate the debate so that metaphysicians may understand the nature and scope of these anti-metaphysical arguments better.

1. Introduction

Metaphysical placement problems are problems with “placing” moral facts (or properties), mathematical facts, or modal facts, and alike, in the world of the senses. Such problems can be provoked by truthmaking questions, i.e., questions of what, if any, makes the sentences and thoughts about a given area of facts true. For example, the question of what makes “Killing babies for fun is wrong” true may be answered by positing a corresponding moral fact. The relevant placement problem asks for an explanation of what it is for there to be such a fact in our world. Non-representationalists (like others in the field, I will use the terms “non-representationalist” and “pragmatist” interchangeably in this paper) reject truthmaking questions to avoid the relevant way of yielding metaphysical placement problems. For example, Blackburn (2010) claims that non-representationalism allows us “to avoid putting any weight on the metaphysical imaginings that it might threaten to engender” when accounting for parts of language and thought. Taking evaluative language and thought for example, he criticizes the posit of “an autonomous realm of values” for being “a metaphysical extra that we inexplicably care about on top of voicing and discussing our more humdrum concerns”. (4–5) Similarly, Michael Williams claims that non-representationalist approaches allow us to avoid “metaphysical worries” about the existence of moral properties and their character. (2013: 130) Huw Price also contends that metaethical expressivism (which is a paradigmatic form of non-representationalism)

“sidesteps the metaphysical conundrums that trouble her representationalist opponents”. (2013: 157) Moreover, non-representationalists sometimes appear to make the stronger claim that their approach allows us to not just sidestep some metaphysical problems but avoid metaphysics altogether when explaining language. For example, Blackburn claims that given non-representationalism, “metaphysics bows out of the picture”. (2015: 850) Price and Macarthur (2007) contend, “[...] our interest is in highlighting the view that begins with such an anthropological concern, without treating it as a stepping-stone to metaphysics.” (6)

However, these non-representationalist arguments do not provide the same motivation for rejecting “truthmaking and metaphysics more broadly” (Simpson 2021). In other words, they accuse truthmaking questions for different reasons. In this paper, I represent and examine three anti-metaphysical arguments in the literature, provided by paradigmatic non-representationalists including Simon Blackburn and Huw Price (which are also the main theorists that Simpson focuses on), discerning two different motivations behind these anti-metaphysical arguments. Then, I argue that the second motivation is more justified than the first. The aim is to elucidate the debate so that metaphysicians may understand the nature and scope of these anti-metaphysical arguments better.

2. The First Anti-Metaphysical Argument

Non-representationalists dismiss truthmaking questions in two typical ways. One is asserting that we should replace truthmaking questions with questions of a more pragmatic character which do not yield placement problems. The other is asserting that, against the background of deflationism about truth, truthmaking questions are either trivial and unproductive or first-order and so non-metaphysical. The two styles of arguments are independent of each other, and we shall discuss each in turn.

Let us consider two different arguments of the first style. They assume differently about the purpose of asking truthmaking questions.

First, consider Blackburn’s (2015) argument that a non-representationalist approach makes truthmaking questions avoidable, and that they should

indeed be avoided for wrongly conceiving of the object of enquiry, i.e., a given subject matter of language and thought. Blackburn contrasts a non-representationalist approach with what he calls a truth-theoretic approach to a given subject matter of language and thought. The truth-theoretic approach asks what the truthmakers of the relevant language and thought are in nature. Blackburn takes it as assuming the relevant subject matter such as “values, norms, natural laws, alternative possibilities, numbers, and others in the same spirit” to be substances that we can put “under a microscope or on a petri dish or in a retort” and “break apart to see what they are made up of”. (850) He claims that, in contrast, a pragmatic approach takes the relevant subject matters to be “categories with which we think”. (850) Accordingly, this latter approach claims to replace the questions about the subject matter itself (cast in truthmaking terms) with questions about what it is to speak or think about that subject matter. In the latter approach, Blackburn contends, “metaphysics bows out of the picture”, and “uses and purposes take their place”. (850)

There, Blackburn didn’t elaborate on what these categories of thought are or how the shift of focus from the language itself to the use of it keeps metaphysics out of the picture. Presumably, he would say that a category of thought is itself not something we can break apart and see what it is made up of, either. It is just another category with which we think. Moreover, similar things can be said for purposes and uses. That is, even if we may have no idea of what categories of thought or uses and purposes are, it will be sufficient for our purposes to have an idea of what it is to think about them.

However, my worry is that similar concerns will require us of an idea of what it is to think about metaphysics, too. But if this is the case, then Blackburn’s argument would either fail to dismiss metaphysics at all or have dismissed other subject matters in the same way it dismisses metaphysics. My diagnosis is that demarcating between the two approaches in terms of their targeting at different aspects of language and thought is not an appealing strategy. Instead, we should concentrate on understanding how they examine the same facets of language and thought differently. Moreover, we should select the relevant facets depending on what the non-representationalist approach targets at.

3. The Second Anti-Metaphysical Argument

Now, let us consider the second argument of the first style. Different from Blackburn, Price (2004) claims that the truth-condition theory of language and thought is not misconceived yet incomplete, in that it doesn't tell us, e.g., *how there comes to be* a thought to the effect that snow is white. In contrast, non-representationalism fills the gap by providing a pragmatic account of *how humans have to be and what they have to do to count as possessing and employing such a thought*. That is, non-representationalism models not languages themselves but instead particular language users and their particular speech acts. In other words, we may say that it studies neither sentences nor statements, but utterances. However, that the truth-condition theory didn't do this doesn't mean that it can't do this. As Price and Macarthur (2007) correctly point out, even "a metaphysician, too, may begin her enquiry with a more-or-less anthropological concern to account for certain aspects of human linguistic behaviour'. They think what distinguishes a non-representationalist approach from the truthmaking approach is their different ways of accounting for the same issues. That is, the non-representationalist account avoids metaphysics in its accounts whereas its rival doesn't. More precisely, non-representationalism does so by rejecting the semantic or representationalist assumption made in the truthmaking approach. On this assumption, linguistic expressions, concepts, propositions, and thoughts stand in a *substantive* semantic relation such as reference and representation to objects, properties, or states of affairs in the world. (5-6) This doesn't mean that non-representationalism rejects any sort of word-world relations. Rather, it means that it takes semantic relations as *theoretically insignificant*. As Price says elsewhere, according to non-representationalism, "there is no additional semantics explanandum, and no distinctively metaphysical question'. (2011: 14)

Price and Macarthur's argument assumes that doing explanation by itself doesn't amount to doing metaphysics, as long as it gives no substantive role to general semantic relations. Price (2004) explains further how non-representationalism avoids placement problems when explaining the use conditions of a sentence. He says that, for example, when explaining what a speaker must know in order to use the utterance "Snow is white" correctly, non-representationalism appeals to the speaker's judgements about colours, not to colours themselves. And because of its rejection of substantive semantic

relations between judgements about colours and colours themselves, the appeal to the former doesn't automatically extend to the latter. One might object that this explanation looks exactly like Blackburn's (2015), in that it also relies on a distinction between language and thought about a subject matter on the one hand, and the subject matter itself on the other. However, notice that we have made clear one crucial difference between the two accounts that Blackburn didn't make clear, that is, non-representationalism doesn't appeal to substantive semantic relations.

There is an alternative way of making sense of the distinction, which allows us to uniformly talk about language and concepts on both sides. Price (2004) introduces it by way of introducing Dummett's (1975) distinction between two conceptions of the task of a theory of meaning. One conception is that a theory of meaning seeks to interpret the target language for "someone who already has the concepts required", whereas the other claims it is to "explain the concepts expressed by primitive terms of the language". (1975: 102) To illustrate, Price quotes McDowell's (1891) claim, which says that on the ground of deflationism about truth, the first conception assigns to a theory of meaning the task of producing biconditionals of the form " s is true iff p ", where p is a sentence of the metalanguage used to show what s , a sentence of the object language, can be used to assert. Notice that such a task is possible only if the conceptual resources of the metalanguage are at least as rich as those of the object language (this will be useful when we defend the third argument). (2004: 190) Price thinks adopting the first conception allows non-representationalists to account for the target sentences without invoking substantive semantic relations.

4. The Third Anti-Metaphysical Argument

Certainly, the sceptic may keep pressing on this: what prevents judgements about colours, say, from being metaphysical judgments? The third argument that we consider addresses this question and hence complements the second. It says that deflationism about truth entails that truthmaking questions are first-order and so non-metaphysical, if not entirely trivial and unproductive. Since contemporary non-representationalists generally adopt deflationism about truth, truthmaking questions will not yield placement problems for

them. The idea is that deflationism about truth provides a trivial answer to any questions of the form “what makes $\langle p \rangle$ true’, namely the answer that what makes $\langle p \rangle$ true is just p . As Blackburn (2012) contends, “Nothing else needs to be said, wearing allegedly metaphysical hats, or allegedly scientific hats.” (195) However, the more crucial point lies in his next claim, according to which if one continues to ask what makes it the case that p —what makes honesty good, for instance—this further question will cease to be about the nature of goodness but becomes instead a first-order question that requires us to draw “whatever connections that impress us between honesty and other things such as respect, integrity, or happiness”. (195-6) That is, although truthmaking questions in this derived sense—i.e., questions about what makes it the case that p —are well-motivated, they are only questions within a specific area such as ethics. For instance, a statement that explains the truth of an ethical statement is itself a first-order ethical statement (also see Blackburn 1998: 294–6). Similarly, Price argues that “what makes it true that snow is white” is “a question to be answered in terms of the physics of ice and light, not in terms of the metaphysics of facts and states of affairs”. (2011: 14)

This last argument assumes that metaphysics of facts and states of affairs is not a first-order matter. To clarify the issue, a distinction between metaphysics *simpliciter* and metaphysics of science may be helpful here. While we may agree with Simpson (2021: 3178) that some first-order issues do seem to have a metaphysical subject matter, e.g., negative truths and facts, truths about logic, parthood, time, causation, free will, induction, and so on, such issues belong to the so-called metaphysics of science. Whereas, what non-representationalists object to is what might be called metaphysics *simpliciter*. It is in this sense of metaphysics that they claim to reject metaphysics “altogether”.

References

Blackburn, S. (1998) *Ruling passions*. Oxford: Oxford University Press.

Blackburn, S. (2010) "The Steps from Doing to Saying", *Proceedings of the Aristotelian Society*, 110, 1–13. Reprinted in his *Practical Tortoise Raising: And Other Philosophical Essays*. Oxford University Press, 169–180.

Blackburn, S. (2012) "Some Remarks about Minimalism", in: C. Wright & A. Coliva (eds.), *Mind, meaning, and knowledge: Themes from the philosophy of Crispin Wright*. Oxford: Oxford University Press, 195–210.

- Blackburn, S. (2015) "Blessed are the Peacemakers", *Philosophical Studies*, 172 (4), 843–853.
- Dummett, M. A. E. (1975) "What is a Theory of Meaning?", in: S. Guttenplan (ed.), *Mind and Language*. Oxford University Press.
- McDowell, J. (1981) "Anti-Realism and the Epistemology of Understanding", in: Bouveresse, J. & Parret, H. (eds.), *Meaning and Understanding*. Berlin: W. de Gruyter, 225–248.
- Price, H. (2004) "Making Sense of Wittgenstein's Linguistic Pluralism", in: Kölbel, M. & Weiss, B. (eds.), *Wittgenstein's lasting significance*. New York: Routledge, 179–206.
- Price, H. & Macarthur, D. (2007) "Pragmatism, Quasi-realism, and the Global Challenge", in: Misak, C. (ed.), *New Pragmatists*. Oxford: Oxford University Press, 91–121.
- Price, H. (2011) *Naturalism without Mirrors*. Oxford: Oxford University Press.
- Price, H. (ed.) (2013) *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.
- Simpson, M. (2021) "Deflationism and Truthmaking", *Synthese* 198, 3157–3181.
- Williams, M. (2013) "How Pragmatists can be Local Expressivists", in: Price, H. (ed.), *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press, 113–128.

Exploring the Empty Sign: Meillassoux's Speculative Materialism and Wittgenstein's *Tractatus*

Jakub Mácha (Brno, Czech Republic)

Abstract

While defending his speculative materialism, Quentin Meillassoux has accused many thinkers of a “correlationism” in which the cognizing subject is incapable of reaching the reality that is indifferent to human thought. He cites Wittgenstein's *Tractatus* and Heidegger's fundamental ontology as prime examples of strong correlationism. In this article, I argue that Meillassoux's speculative materialism closely resembles Wittgenstein's *Tractatus*. I argue that Wittgenstein's account of names and objects is very close to Meillassoux's theory of the empty sign, devoid of meaning and capable of identical iteration. This, he claims, is the distinguishing feature of mathematical discourse, which is capable of accessing mind-independent reality. If this is so, then either some correlations are capable of accessing this reality or, more importantly, the notion of correlation needs to be specified more precisely. At the same time, I will argue that Meillassoux's derivation of the empty sign can be seen as complementary to the *Tractatus*, since Meillassoux argues that a sign is thinkable without any meaning. In the final part of the paper, I will evaluate Meillassoux's argument for the facticity of every correlation, and argue that Wittgenstein's *Tractatus*, under certain interpretations, can escape this criticism.

1. Meillassoux's derivation of the sign devoid of meaning

Quentin Meillassoux has argued that mathematical languages are able to represent reality independently of thought, without resorting to the principle of sufficient reason and correlation. The distinguishing feature of such languages, he argues, is their ability to conceive of the empty sign – a sign devoid of meaning. This minimal requirement is what allows them to capture the absolute contingency of entities. Several authors, most notably Livingston (2012) and van Gerven Oei (2014), have argued that Meillassoux is mistaken in thinking that empty signs mark the distinctiveness of formal languages. They, drawing on insights from Derrida's deconstruction, suggest that empty signs mark formal aspects of any language. This is an apt criticism. Besides, this point brings Meillassoux's account of empty signs closer to Wittgenstein's *Tractatus*. Wittgenstein differentiates between everyday language and sign-language (TLP: 3.323, 3.325). The purpose of sign-language is to reveal the inherent logic of everyday language. Everyday language is “logically completely in order” (TLP: 5.5563).

We need to distinguish two claims: first, that it is possible to think the empty sign; and second, that the capacity to think the empty sign enables access to

mind-independent reality. Meillassoux provides a detailed argument for the former claim but only briefly sketches the latter. I will argue that Meillassoux's empty sign corresponds to Wittgenstein's concept of a sign (a perceptible mark) that is taken independently of its symbol or meaning. Therefore, the capacity to think the empty sign, within Wittgenstein's framework, means the ability to take a sign independently of its symbol. Wittgenstein does not provide an argument for the existence of this capacity. Meillassoux's derivation of the empty sign can, therefore, be seen as supporting Wittgenstein's distinction between sign and symbol.

To understand the concept of a sign devoid of meaning, consider the following observation: ZF set theory is grounded in axioms that dictate what can be done with sets, but it lacks an explicit definition of what a set is. Instead, the axioms serve as rules for how signs representing sets can be combined. These signs lack any preassigned or precorrelated reference, making them empty or devoid of meaning.

The term "devoid of meaning" is somewhat misleading and implies a referential theory of meaning; better would be "devoid of explicit definition." This problem may be due to the translation of the French "sens" as "meaning." "Sens" can also mean "sense" or "direction" (see IRR: 195, fn. 38). Meillassoux argues that a sign is empty because it does not appear as the *definiendum* in a definition of the form *definiendum = definiens*. However, empty signs can still appear in normative rules, such as the ZF axioms, that specify what can be done with those signs. As Livingston (2012) states, Meillassoux believes that defining a sign contextually, known as an implicit definition, does constitute equipping it with meaning. However, for Meillassoux, the sole way to imbue a sign with meaning is by correlating it to something outside of the system – an external referent. This is an idiosyncratic use of the term "meaning". It would render many parts of natural languages devoid of meaning, such as adjectives, adverbs, verbs, and terms for abstract concepts. Therefore, an empty sign cannot be considered the distinctive feature of formal languages.

Before delving into the nature of the empty sign, it is important to introduce Meillassoux's distinction between base-signs and operator-signs. Base-signs consist of constants and variables and are supposed to be empty, whereas operator-signs typically involve logical or mathematical connectives and allow

the formulation of rules for what can be done with base-signs. However, Wittgenstein argues that these operator-signs must be eliminated from a logically perspicuous language, and he takes this to be *the* fundamental insight of the *Tractatus* (TLP: 4.0312). Wittgenstein presents a complex argument in the *Tractatus* that reduces all logical connectives to the generalized Sheffer stroke and introduces the general form of propositions, where the logical operation is captured by its formal properties. This argument can be adopted into Meillassoux's theory to enrich and strengthen his speculative materialism.

Let us now return to base-signs, which are devoid of explicit definition. Meillassoux claims that base-signs are arbitrary, meaning that any perceptible mark can serve as a base-sign. The visual or auditory properties that represent a concrete sign are arbitrary and unrelated to the meaning or function of the sign. Meillassoux argues that this radical arbitrariness allows the sign to capture the contingency of every entity.

Meillassoux attempts to demonstrate the possibility of the empty sign, and reflects on how it can be recognized as the same sign across numerous instances, given its arbitrary nature. He proposes a unique solution: The empty sign must be capable of potentially unlimited reproduction, and while each instance may differ slightly (as, for instance, in handwriting), it must still be recognizable as an instance of the same sign.

Meillassoux investigates the difference between repetition of the same ornamental pattern and iteration of the same sign (and similarly, for the auditory modality, between repetition of the same tone and iteration of the same acoustic signal). Although the perceptible appearance of these two series is the same, Meillassoux introduces the following terminology to distinguish them: A series of ornamental patterns is called *repetition*, while a series of signs is called *iteration*. The question then arises: What distinguishes repetition from iteration?

Meillassoux draws on Bergson's idea that repetition produces a differential effect that is not based on any distinguishability between patterns or tones. The same tone is heard differently when repeated. This effect is called the *threnody* effect, and although Bergson believed it only occurs in the auditory modality, Meillassoux argues that it occurs in the visual one, too. He calls this

the *frieze* effect and provides architectural examples. If a sign is to be recognized repeatedly, this differential effect must be negated.

Meillassoux posits a connection between the iterability of a sign, its arbitrariness, and the contingency of every entity. According to him, the contingency of a thing is iterable identically from mark to mark without any differential effect of repetition (IRR: 180). This connection between contingency and iterability indicates that a sign must be seen as arbitrary, i.e., replaceable by any other shape if it is capable of unlimited iteration. In other words, any visual or auditory appearance is contingent, and to see a thing as an iterable sign is to see this thing in its contingency.

To understand how a thing can be perceived in its contingency, Meillassoux distinguishes between two modes of perception: ordinary and semiotic. In ordinary perception, we first grasp the thing through its properties and only secondarily consider its contingency/its facticity, i.e., the fact that it is given to us as what it is (IRR: 182). On the other hand, in semiotic perception, we begin with the thing's facticity, which surrounds it like a "diffuse aura" (ibid.). Perceiving a thing in this way allows it to be iterated, meaning it can escape the differential effect of space-time. This does not mean that empirical properties of the thing are disregarded. A visual mark perceived as a sign retains its empirical properties, but its facticity, its occurrence within a fact, takes precedence over its physical properties.

Meillassoux proposes a way to recognize a thing as a sign without invoking meaning or reference: by perceiving it in its facticity. In ordinary perception, we first grasp a thing through its properties and then consider its contingency. In contrast, semiotic perception begins with a thing's facticity as its primary aspect, while its physical properties are secondary. Meillassoux has yet to provide a comprehensive argument for how such a sign can access mind-independent reality. Meillassoux acknowledges this when he concludes an earlier draft of his article "Iteration, Reiteration, Repetition" with the statement: "But we have not at all shown that the empty sign allows, in turn, the description of a world independent of thought." (2012: 37) In the next section, we will explore the Tractarian account of names and variables, which aligns with Meillassoux's proposed view of semiotic perception.

2. Names in the *Tractatus* are empty signs

I argue that Tractarian names and variables capture the essence of Meillassoux's notion of empty signs. Wittgenstein's logical atomism highlights the primacy of facts over things, as stated at the beginning of the *Tractatus*: "The world is the totality of facts, not of things" (TLP: 1.1). Objects do not have material properties, only formal ones, which show how objects combine into atomic facts (TLP: 2.011). Meillassoux describes Galilean materialism as the view that matter has no qualities and is therefore "entirely mathematically describable" (2010: 6). As signs are objects like everything else, they are iterable by default. Thus, any fact can assume the role of an empty sign.

To understand the Tractarian account of signs, we can turn to Wittgenstein's clear definition: "The sign is the part of the symbol perceptible by the senses" (TLP: 3.32). A symbol is a sign with meaning, whereas a sign considered in isolation, without its meaning, is abstracted from a symbol and, therefore, devoid of meaning. Wittgenstein's terminology of "sign" and "symbol" corresponds to Meillassoux's distinction between a "sign devoid of meaning" and a "sign provided with meaning" (IRR: 182).

To clarify, Tractarian names are devoid of explicit definition, which is what Meillassoux means by "devoid of meaning." Wittgenstein confirms this at several points throughout the *Tractatus* and his later work. For example, he says that "objects can only be *named*" and "I can only speak *about* them: I cannot *put them into words*" (TLP: 3.22). While the meaning of a name cannot be put into words, it must be explained to us for us to understand it (TLP: 4.026). This is true in ordinary perception, where names and objects are considered in isolation from their roles in propositional signs and facts. However, in semiotic perception, which is the default mode of apprehension in the *Tractatus*, no such explanation is needed. Wittgenstein emphasizes this in a later remark:

3.33 In logical syntax the meaning of a sign should never play a role. It must be possible to establish logical syntax without mentioning the *meaning* of a sign.

The remark emphasizes that in logical syntax, signs are defined based on their combinatorial properties rather than explicit referential definitions. As a result, these signs are devoid of meaning.

I will now argue that, according to Meillassoux, these signs are arbitrary. Wittgenstein believed that a notation is essentially arbitrary (TLP: 3.342), as the way it is produced determines its arbitrary character. What is necessary is its capacity to signify, as “the real name of an object was what all symbols that signified it had in common” (TLP: 3.3411). This means that a name does not necessarily have to signify the object it does. There cannot be a pictorial relationship between names and objects, even though they are part of the overall picture theory of meaning.

3. Names as variables

To put my main point more clearly and succinctly: Tractarian names, when viewed as variables, are equivalent to Meillassoux's signs devoid of meaning. This is because the arbitrary nature of the notation used to signify objects means that the name can be substituted with any object, and its formal properties, rather than its referential ones, are what matter.

Wittgenstein's account of variables is peculiar in that variables must always occur within a proposition. There are no individual variables that refer to objects, only propositional variables whose values are the propositions that contain the expression (TLP: 3.313). It is incorrect to say that x ranges over objects a , b , c , etc. Rather, the propositional variable “There is an x ” ranges over propositions like “There is an a ,” “There is a b ,” and so on. Wittgenstein's crucial point, and this is relevant to my argument, is that:

5.526 We can describe the world completely by means of fully generalized propositions, i.e. without first correlating any name with a particular object.

This is the most explicit expression of Wittgenstein's anticorrelationism in the *Tractatus*. Wittgenstein also refers to such description as “impersonal representation of the world” in his *Notebooks* (NB: 20). This complete representation is achieved by turning all simple signs into variables, which are

signs without explicit referential definition: “All the signs in [a proposition] that have arbitrarily determined meanings are turned into variables” (TLP: 3.315).

Atomic propositions are composed solely of names, which are then converted into variables, resulting in fully generalized propositions that are composed entirely of variables. Wittgenstein gives an example of such a proposition:

$$(\exists x, \Phi). \Phi x \quad (\text{TLP: 5.5261})$$

This has the logical form of second-order predicate calculus, as it quantifies over predicates (Φ). However, adopting such a logic raises philosophical issues, mainly due to its incompleteness. While I cannot discuss these issues here, they must be addressed to support the argument that the world is fully describable through fully generalized propositions. Wittgenstein grapples with the tension between the contingency of fully generalized propositions and their necessity as formal properties of logical notation in a remark from his *Notebooks*:

What the completely general propositions describe are indeed in a certain sense structural properties of the world. Nevertheless these propositions can still be true or false. Even after they have meaning, the world still has that range [*Spielraum*]. (NB: 20, translation modified)

The German word *Spielraum* captures the idea of contingency (*Spiel*) within a general framework (*Raum*). Wittgenstein explains what this *Spielraum* is in the *Tractatus*: “And the range that the totality of elementary propositions leaves open for its construction is exactly the same as that which is delimited by entirely general propositions” (TLP: 5.5262). The formal properties of the logical notation correspond to the structural properties of the world. Wittgenstein sees the formal properties of the world as contingent, allowing generalized propositions to be true or false. This type of contingency is distinct from *Spielraum*. (Bradley (1992) envisages this kind of higher-order possibility in the *Tractatus*, though his argument and textual evidence differ from my approach.) Wittgenstein’s acknowledgment of this radical contingency – which is given little emphasis in the *Tractatus* – enables him to avoid Meillassoux’s critique, which I will present shortly.

4. Meillassoux's critique of the *Tractatus*

Meillassoux criticizes what he calls strong correlationism and later subjectivism, which he argues is endorsed in *Tractatus*. He puts forward the argument of *correlational facticity*, which highlights the “*thinkable nonnecessity of correlation*” (IRR: 135). Meillassoux agrees with the strong correlationist that there is no reason for the correlation itself; but he is concerned that the strong correlationist takes the correlation as the *arche-fact*, meaning a fact that cannot be conceived as different from what it is or as not existing, but whose necessity cannot be demonstrated (ibid.).

How does Meillassoux's argument apply to Wittgenstein's case? Meillassoux quotes several remarks about the mystical from the *Tractatus*, such as 6.522 (cited in AF: 41–42): “There are, indeed, things that cannot be put into words. [...] They are what is mystical.” This wording suggests that there are *things* or *objects* that cannot be named and thus cannot be expressed in language. However, the expression “things” does not occur in the German original, which refers merely to “*Unaussprechliches*” – rendered more directly in Ogden/Ramsey's older translation as: “There is indeed the *inexpressible*. [...] it is the mystical.” This inexpressible/mystical is specified in 6.44 (also quoted by Meillassoux): “It is not *how* things are in the world that is mystical, but *that* it exists.” Again, the word “things” has no counterpart in the German original, and it is misleading to say that there are inexpressible mystical things. Ogden/Ramsey's translation is more accurate: “Not *how* the world is, is the mystical but *that* it is.” Recall that the term “world” has a technical sense in the *Tractatus*: “The world is the totality of facts, not of things” (TLP: 1.1). (M. Beaney's recent translation accurately renders 6.44 and 6.522, without implying that Wittgenstein was committed to the existence of mystical things.) It is thus misleading to say that there are inexpressible mystical things. What is inexpressible is that there is the totality of facts. In *Time Without Becoming*, Meillassoux uses slightly different wording: “The *Tractatus Logico-Philosophicus* [...] designates as ‘mystical’ the mere fact that there is a consistent world” (2014: 12), which is a plausible paraphrase of Wittgenstein's words. However, why speak here of “consistency”? The mere fact that there is a consistent world does not imply anything about its correlation with language or thought.

Wittgenstein says quite explicitly that the world can be changed: “If good or bad willing changes the world, it can only change the limits of the world, not the facts; not what can be expressed by means of language” (TLP: 6.43). The world could have been constituted by a different totality of facts, even without there being any single fact that differs between that other totality and the one that actually constitutes the world. This means that the world is contingent in Meillassoux’s sense of the term, i.e., it “could have existed otherwise” (IRR: 135). On the other hand, it is mystical whether the world is capable of not being – which is another sense of “contingent” in Meillassoux.

Since a different world is conceivable, it is thus *factual* (again, in Meillassoux’s sense). As we know, the world is fully describable by means of fully generalized propositions. We can conceive a different collection of such propositions. That means that different fully generalized propositions will be true or false. Such a different world is thinkable outside any correlation between names and objects. Of course, within our world and language we cannot express factual statements about the other world because we lack its language. This language has a different collection of names or names with different internal/combinatorial properties.

Hence, the world is contingent and factual. This contingency and facticity are of a more fundamental order than the contingency and facticity of empirical facts (see IRR: 135). This is how the *Tractatus* can accommodate correlational facticity. And so the argument presented above can be employed to support (not against) the general outlook espoused in the *Tractatus*.

References

Bradley, R. (1992) *The Nature of All Being. A Study of Wittgenstein’s Modal Atomism*. Oxford: Oxford University Press.

Livingston, P. M. (2012) “Realism and the Infinite”, Draft version. <https://philpapers.org/archive/LIVRAT.pdf> (Access: February 16, 2024).

Meillassoux, Q. (2006) *After Finitude*. Trans. by R. Brassier. London: Continuum, 2008. Abbreviated as AF.

Meillassoux, Q. (2012) “Iteration, Reiteration, Repetition: A Speculative Analysis of the Meaningless Sign”. Draft version. Translated by R. Mackay. <https://s3.amazonaws.com/arena-attachments/886529/539a4b4a8c213179c159eefc04a28947.pdf> (Access: February 15, 2024).

Meillassoux, Q. (2014) *Time Without Becoming*, Fano: Mimesis International.

Meillassoux, Q. (2016) "Iteration, Reiteration, Repetition: A Speculative Analysis of the Sign Devoid of Meaning", in: A. Avanesian, S. Malik (eds.), *Genealogies of Speculation: Materialism and Subjectivity since Structuralism*. London: Bloomsbury, 117–197. Abbreviated as IRR.

van Gerven Oei, V. (2014) "By Any Language Necessary: Quentin Meillassoux and the Question Concerning Signification in Philosophy", *Quaderna 2*. <https://quaderna.org/wp-content/uploads/2014/05/by-any-language-necessary-quentin-meillassoux-and-the-question-concerning-signification-in-philosophy.pdf> (Access: February 15, 2024).

Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, London, Paul Kegan, 1922. Abbreviated as TLP. I use both the Ogden/Ramsey and the Pears/McGuinness translations, and in some cases combine or amend them. Also M. Beaney's recent translation (2020) has been considered.

On Granularity Matching and Scale Mapping: A Novel Challenge for Multiple Realization

Matteo Mattersberger (Vienna, Austria)

Abstract

The present paper aims to discuss a challenge for assessing the multiple realization (MR) thesis, which states that no (kind of) mental state can be identified with any (kind of) physical state, as mental states can be brought about by various physical substrates (i.e., multiply realized). Following this claim, various alleged examples of MR have been proposed in the philosophical literature. However, as Bechtel and Mundale argue, many of these examples suffer from specifying the mental states in question at a very coarse level, while specifying their physical substrates at a very fine level (Bechtel & Mundale 1999), thus artificially creating the appearance of MR. Based on their critique, I propose the Granularity Matching Criterion (GMC) as a benchmark for evaluating MR claims, which states that for a claim of MR to hold, the granularity of the mental state's description in question must match that of the physical state's description. However, this new benchmark, as will be attempted to show, suffers from our inability to systematically assess whether mental and physical states are adequately matched in granularity, a problem to which I refer as the "scale mapping problem". This inability to systematically map the granularities of physical and mental state characterizations poses, so will be argued, a significant obstacle to validating MR claims, an obstacle which needs to be overcome to assess alleged cases of MR.

The empirical thesis

Functionalism, the thesis that mental states are best individuated by the causal role they play in the larger cognitive architecture they are part of, is a common assumption in the fields of cognitive science and philosophy of mind (Sterelny 1990). This thesis about the nature of mental states was put forward to account for the notion of multiple realizability of mental states, which has been largely accepted as orthodoxy in some areas of philosophy. The argument, going back to Hilary Putnam, states that the same mental kinds can be brought about by a myriad of different physical substrates, which means that any kind of mental state cannot be identified with any specific kind of physical substrate (Putnam 1967). Importantly, Putnam takes multiple realization (henceforth MR) to be not merely a logical possibility, but rather an empirically testable aspect of reality, i.e., he takes mental states to be actually multiply realized in different species, as opposed to this MR being a mere logical possibility. Thus, following this thesis, through careful empirical observation we should be able to find examples of differently realized mental state kinds.

In the literature following Putnam's argument, this empirical thesis has been developed into concrete case studies, with proponents of the MR thesis bringing forward various examples of mental states being, allegedly, realized by distinct neural kinds (Polger & Shapiro 2016). To give a concrete example, Aizawa and Gillett propose that the mental kind *long-term memory consolidation* is multiply realized, as the neurochemical process underlying this mental state involves different amino acids in different species of animal, namely *Drosophila*, *Aplysia*, and mice (Aizawa & Gillett 2009). They characterize this process as follows:

During electrical activity of neurons, cells come to have increased levels of a secondary messenger, cyclic adenosine monophosphate (cAMP). cAMP molecules bind to molecules of protein kinase A (PKA) [...]. When cAMP binds to the PKA regulatory subunits, the entire molecule dissociates leaving two free PKA catalytic subunits in the cytosol. As the concentration of PKA catalytic subunits increases in the cytosol, the concentration also increases in the neuronal cell nuclei, where they add phosphate groups to various other molecules farther along in the biochemical cascade. Eventually, this biochemical cascade of reactions leads to the synthesis of more proteins that effect structural changes in the neuronal synapses. (Aizawa & Gillett 2009: 198).

Whereas others have argued that this process is underlying memory potentiation across species (Bickle 2003), Aizawa and Gillett claim (in response to Bickle) that differences in the biochemical details laid out above justify the claim of long-term memory potentiation being multiply realized, as "PKA corresponds to distinct amino acid sequences in the different species of organisms [under consideration]" (Aizawa and Gillett 2009: 198). In short, they argue that the involvement of different amino acids in the described biochemical process across species suffices for the MR of the mental state it brings about. Let us examine this claim more closely, as such an assessment will serve as a first step in identifying more deeply rooted problems in the MR literature.

Granularity matching

To begin said examination, I shall first attempt an intuition pump by asking you to imagine the following scenario:

Following Putnam's empirical thesis, I give you an (alleged) concrete example of MR. This example, I claim, shows that mental states are not only differently realized in different species, but within the same species, even within a single brain. The example is this: The mental kind *perception* is multiply realized within the human brain, as different instances of it are realized by spatially and functionally distinct and independent areas within the human brain. Some instances of *perception* are realized by area V1 in the occipital lobe, others by area A1 in the temporal lobe, and others by S1 in the rostral parietal lobe. Thus, there are demonstrable examples of MR.

Even though at first glance this may seem like an interesting observation for those not too well versed in neuroscience, I imagine this assessment would dwindle quickly upon learning that area V1 is involved in vision processing, area A1 in auditory processing, and area S1 in somatosensory processing (Bear et al. 2020). Equipped with this knowledge, one would intuitively be justified in objecting that the example above is not a case of MR at all. Rather, it merely has the superficial appearance of such because the description of the mental kind involved, namely perception, is far too abstract a characterization compared to the characterization of the neurophysiological kinds in question. Whereas the latter denote modality-specific modes of perceptual processing, the mental kind characterization *perception* generalizes over these different modalities, thus ignoring important differences. I shall refer to such cases as cases of "pseudo Multiple Realization" (henceforth pseudo-MR).

The problem of pseudo-MR was first identified by Bechtel and Mundale, who (although using different nomenclature) argue that such cases are abundant in philosophical MR literature (Bechtel & Mundale 1999). These mismatches arise when mental and physical states are characterized at different levels of granularity (henceforth LOG). For instance, the mental state of recalling a childhood field trip might be categorized with varying specificity, as memory, long-term memory, episodic declarative long-term memory, etc. depending on the chosen LOG (Sternberg & Sternberg 2016). Similarly, a neural state might be identified with activation of certain broad brain regions, networks within

that region, individual neurons interacting, specific biochemical reactions, etc. A state kind's LOG depends on the applied degree of abstraction from differences between individual states. Therefore, whether or not two states can be subsumed under a shared kind (partially) depends on the chosen LOG used for characterizing said kind. The states *knowing how to ride a bike* and *remembering last year's Christmas* could be subsumed under the shared kind "memory" or be considered to belong to the distinct kinds *procedural memory* and *declarative memory* (Sternberg & Sternberg 2016), depending on the applied LOG. Following this line of reasoning, granularity mismatch occurs when one arbitrarily abstracts away differences between mental states when subsuming them under a shared kind, while emphasizing the same kind of differences at the physical level to argue that the physical states in question belong to different kinds.

Bechtel and Mundale employ this line of reasoning to argue that mental states are, despite many claims to the contrary, not multiply realized, and that seeming cases of MR can be explained away by the foregoing analysis (Bechtel & Mundale 1999). For the purposes of the present paper, I wish to remain agnostic on this matter. However, independent of one's inclination to follow Bechtel and Mundale in their rejection of MR, their analysis gives us a benchmark against which to measure alleged cases of MR to assess their validity. Based on their analysis, I want to propose the following criterion:

Granularity Matching Criterion (GMC): An example of the MR of a mental state by more than one physical state must employ either the same or a higher LOG for characterizing the mental state compared to the LOG employed to characterize the physical state.

The reasoning for formulating the GMC so that the level of granularity for characterizing the mental state can be finer than the one characterizing the physical state is that, if anything, such an example would make for an even stronger case of MR, e.g. if the exact same very narrowly defined mental state can be instantiated by a brain and a digital computer. Adherence to this criterion when proposing a case of MR ensures avoiding the problems of pseudo-MR due to granularity mismatch discussed in the foregoing paragraphs.

Having established this criterion for assessing potential cases of MR, recall now the previously discussed case study proposed by Aizawa and Gillett, who claim that long-term memory consolidation is multiply realized in different species (Aizawa & Gillett 2009). Applying the GMC to this case reveals it to be deeply problematic as an example of MR, as there seems to be a gross mismatch between the very coarse characterization of the mental states in question and the very fine characterization of their physical substrate. The authors explicitly spell out this coarse-grained characterization (although without deeming it problematic) when they state the following:

[T]here is a single cognitive process of memory consolidation that may be found in *Aplysia*, *Drosophila*, and mice [...]. [B]y ‘memory consolidation’ we shall understand any cognitive process that enables memories to persist for hours, days, and weeks, rather than merely seconds or minutes. Roughly speaking, memory consolidation is the process of transforming short term memories into long term memories. (Aizawa & Gillett 2009: 197).

Contrary to this very general characterization of memory consolidation, which merely states its functional role in the broadest possible term, the authors characterize the physical underpinnings of the mental state in question on a very fine level, namely that of the biochemical brain processes specified earlier, and claim that minute differences in chemical make-up at this level suffice for a genuine case of MR.

If one is convinced by the foregoing arguments regarding granularity mismatch, then this example seems to be a paradigmatic case of it: Small differences at the mental level are glossed over to subsume the mental states in question under a shared kind, while the characterization of the physical substrate accentuates such small differences between individual physical states. However, it seems eminently plausible that one could employ an equally coarse LOG for characterizing the physical states in question, much like Aizawa and Gillett do for the relevant mental states. By doing so, one could subsume the physical states realizing memory consolidation in all three species under a single kind, glossing over the minor differences in amino acids between the physical substrates and focusing instead on the shared biochemical process.

Scale mapping

So far, we have merely followed Bechtel and Mundale in their critique of MR. However, the problem of the empirical thesis lies deeper than at a mere failure to adhere to the GMC. To see that this is the case, simply imagine, to continue with the example of memory consolidation, that Aizawa and Gillett based their claim on the MR of long-term memory potentiation on a difference in the neurophysiological pathways involved in realizing memory consolidation instead of a difference at the biochemical level. Would this match the LOG of the mental characterization? What if entirely distinct brain areas were involved? One could also go the other way and attempt to adjust the LOG of the mental state characterization in order to achieve the granularity matching demanded by the GMC. Instead of giving the highly general characterization *long-term memory consolidation*, one could give the more specific *characterization implicit (non-declarative) memory consolidation* or the even finer grained characterization *procedural memory consolidation*. Would any of these be sufficiently close in granularity to the LOG of the given substrate characterization to fulfill the GMC? It seems entirely unclear. This unclarity, however, is highly problematic for the MR discussion. While it seems possible to judge cases of mismatched granularity in examples with extreme disparities in the levels of characterization granularity (such as the mock example given above) by mere intuition, this intuition-based strategy seems to break down in cases such as the examples just given. Said strategy of judging whether granularities of characterizations are matched relies on disparities between the two LOGs being, metaphorically speaking, visible with the naked eye, i.e., assessable without relying on some sort of systematized criteria of evaluation. However, it cannot be assumed that this is usually the case. It seems unclear when exactly two state characterizations at different levels are actually matched in granularity and when this is not the case, unless, as just discussed, the difference in LOG is so apparent that mere intuition suffices to recognize it.

Let us put the problem differently: It is possible to give characterizations of the same mental and physical states at different LOGs. It also seems evident that the LOGs within each domain (mental/physical) can be ranked from coarsest to finest, on a within-domain granularity scale. *Memory consolidation* is coarser-grained than *implicit memory consolidation*, which in turn is more coarse-

grained than *implicit procedural memory consolidation*. However, the problem seems to be that it is entirely unclear how the mental and the physical granularity scales map onto each other. In other words, it is entirely unclear for any given characterization on the mental level what the corresponding (in terms of granularity) characterization on the physical level would be. I shall refer to this problem as the problem of “scale mapping”.

Fulfilling the GMC when giving an example of potential MR seems to require first solving the scale mapping problem, as the former requires matching of LOGs across scales, which is impossible without first finding a way to map the LOG of a given state in one domain to the LOG of the corresponding state in the other domain (this requirement does, following what has been said previously, not apply if the mental level characterization is more fine-grained to a degree that it is detectable by the intuition-based strategy).

If one takes the position of Bechtel and Mundale and the newly introduced GMC seriously (which, I have attempted to argue, one should), this inability of scale mapping has far reaching consequences for the debate on MR, as it implies that there is no systematic way (the intuition-based strategy is by its very nature not systematized) to assess potential cases of MR.

Having to rely on the intuition-based strategy alone would mean that to find support for MR (as an empirical thesis), one would have to encounter a case in which the physical substrate of a mental kind is so fundamentally different from any known substrate that it is obvious that the difference would hold even at a very coarse level of granularity, so that it is, to pick up an earlier metaphor, visible by the naked eye (i.e., the intuition-based strategy). Encountering alien life of a completely different biochemical and functional make-up or an advanced artificial intelligence which possess mental states comparable to ours would qualify as such a case. This, therefore, means that sufficient progress in the fields of computer science, engineering, or xenobiology may provide a genuine example of MR, as the MR would in this case be detectable by the intuition-based strategy.

Why then is the intuition-based strategy alone not a sufficient solution for the problem laid out above? The reason lies in its reduced sensitivity to cases of MR. In other words, the strategy, while likely to detect cases of MR that hold at

a coarse LOG at the physical level, is unlikely to detect cases which do not hold with such a coarse-grained physical characterization. It follows that many of the potentially interesting potential examples of MR one can find in comparative neuroscience could not be sensibly discussed by employing the intuition-based strategy alone, as this strategy is blind to all but the most extreme (in respect to LOG disparity) cases of MR. It is entirely possible that there are no examples in (earthly) neurobiology in which the disparity between mental and physical state characterization LOG is extreme enough to allow for the intuition-based strategy, meaning that with our current state of knowledge, we could not adequately assess the empirical MR thesis, a result that goes against the hopes and practices of many philosophers and cognitive scientists.

As I take it that assessing individual cases of MR is an interesting and fruitful endeavor, the foregoing discussion leads to the question how the problem of scale mapping can be solved *in a systematic manner*. In other words, the question is how to judge whether for any given mental state m and physical state p the thesis that the LOG of the characterization of m is equal to the LOG of the characterization of p , so that $G^m = G^p$ (with G being the level of granularity of the given state) holds. Although I have stressed the importance of solving the problem, I must reluctantly admit that I cannot propose a satisfying solution.

Conclusion

In the previous sections, I have pursued two primary argumentative goals.

First, I have attempted to reconstruct and elaborate upon Bechtel and Mundale's critique of common practices in the literature on MR, namely to identify mental kinds characterized at a very coarse LOG while distinguishing physical kinds based on fine grained differences (Bechtel & Mundale 1999). Following these authors, I have argued that when formulating an alleged case of MR, one ought to ensure that the characterization of the realizer state and the realized state share the LOG or that the realizer (i.e., physical) state should be more coarsely characterized, a criterion which I have referred to as the "Granularity Matching Criterion (GMC)". Failing to adhere to the GMC, I have

argued, could lead to cases of pseudo-MR, i.e., cases which superficially resemble cases of genuine MR but fail as such upon closer examination.

Based on this assessment, I have then discussed what it could mean for two state characterizations at different (i.e., mental and physical) levels to share the same degree of granularity. In this discussion I have concluded that the answer to this question is currently unclear because there is no systematized way to map the granularity at one level to the granularity at the other level. It seems like we can only rank the LOGs of characterizations of one and the same state, but not across different states in different domains (mental and physical). Due to this scale mapping problem, it seems impossible to systematically (i.e., beyond the capacities of the intuition-based strategy) ensure that the GMC is fulfilled. I have therefore emphasized the importance of finding a way to map granularities at different levels onto each other in order to systematically assess potential cases of MR.

References

- Aizawa, Kenneth; Gillett, Carl (2009) "The (Multiple) Realization of Psychological And Other Properties In The Sciences", *Mind & Language* 24, 181-208.
- Bear, Mark; Connors, Barry; Paradiso, Michael (2020) *Neuroscience: Exploring the Brain*, Burlington: Jones & Bartlett Learning.
- Bechtel, William; Mundale, Jennifer (1999) "Multiple Realizability Revisited: Linking Cognitive And Neural States", *Philosophy of Science* 66, 175-207.
- Bickle, John (2003) *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Springer Dordrecht.
- Polger, Thomas; Shapiro, Lawrence (2016) *The Multiple Realization Book*, Oxford: Oxford University Press.
- Putnam, Hilary (1967) "Psychological Predicates", *Art, Mind, and Religion*, 1, 37-48.
- Sterelny, Kim (1990) *The Representational Theory of Mind: An Introduction*, Cambridge (MA): Blackwell.
- Sternberg, Robert; Sternberg, Karin (2016) *Cognitive Psychology, 7th edition*, Boston: Cengage Learning.

Kripke and Wittgenstein on Essence, Necessity, and the A Priori

Jacob McDowell (Chicago, USA)

Abstract

I respond to Panu Raatikainen's contribution to the recent volume *Engaging Kripke with Wittgenstein* in order to revisit Wittgenstein's views on such topics as necessity, the a priori, and essence. Raatikainen's chapter, "On the Alleged Incompatibility Between Wittgenstein and Kripke", argues that many of the perceived incompatibilities between the later Wittgenstein and Kripke's views in *Naming and Necessity* are merely apparent, and that some real points of incompatibility are points in Kripke's favour. Against Raatikainen, I argue that many of the incompatibilities Raatikainen takes to be merely alleged are real, and that they speak to deep tensions within Kripke's approach. More broadly, this shows that Wittgenstein's thinking on these issues is not merely of historical interest but bears upon what has become the orthodox stance on necessity, the a priori, and essence in contemporary analytic philosophy.

I begin by analysing Raatikainen's claims that Kripke was not offering a general theory of language and that Wittgenstein's discussion of the Augustinian picture has no bearing here. In section two I lay out what I take to be the core differences between the later Wittgenstein and Kripke on essences and the a priori, differences Raatikainen neglects. Finally, in Section 3, I argue that these differences explain tensions in the Kripkean picture that relate back to Wittgenstein's criticisms of the Augustinian picture.

1. Theories of language and the Augustinian Picture

Saul Kripke famously suggested that proper names and natural kind terms are rigid designators that denote the same thing in all possible worlds where that thing exists, and then both are introduced by initial baptisms or acts of ostension from which the name or term is causally passed down to subsequent users. Users of such a name or term need not associate it with a true description in order to succeed in referring to the kind or entity in question. In the case of a natural kind term, like gold, the discovery that gold is the element with atomic number 79 was the discovery of the essence of gold, the discovery of a necessary a posteriori truth. Earlier reference fixing definitions like 'gold is a yellow, metal' were merely contingent a priori truths. Panu Raatikainen undertakes to show that these views of Kripke's do not constitute a theory to which Wittgenstein or his followers should ought to have any strong objections:

(...) a project of providing the foundation of language is emphatically not what Kripke was aiming at. Therefore, he could have easily agreed with Wittgenstein ... He was simply noticing that the introduction of a new

proper name often involves ostension. There can well already be a lot of language (or language-like thought) in place. (Raatikainen 2024: 13)

Kripke did claim not to be giving any general theory of language and was generally opposed to any reductionist account of meaning or intentionality. But that does not mean that there are Wittgensteinian objections to be made to some of the specific cases or classes of terms Kripke discusses. Whatever the full details of the picture may be, Kripke's discussion of natural kind terms as rigid designators whose reference is established by ostension to a set of stereotypes thereby establishing a metaphysically necessary truth and then subsequently discovered, is already plenty for the Wittgensteinian to object to. This account need not be supposed to generalize to all of language (as Kripke indeed didn't suppose it to), nor need the details be spelt out in reductionist terms or otherwise. The very idea of ostension to an essence that determines what it is to correct to say in all possible scenarios which we discover is, I believe, already nonsense by Wittgensteinian lights. Simply saying that Kripke's theory was not wholly general does not exempt it from Wittgenstein inspired criticisms. I will try to flesh out those criticisms in subsequent sections.

Raatikainen also maintains that Kripke's view is exempt from Wittgenstein's complaints about the Augustinian Picture of language, on the ground that those complaints were directed toward a very specific theory of ostensive definition. "In summary, at least as far as the focus is on "the Augustinian picture" and ostensive definitions (or ostensive teaching), the alleged incompatibility of the views of Kripke and Wittgenstein are merely specious and simply not real." (Raatikainen 2024: 14).

That is, because Kripke is not giving a general theory of language or even a Millian direct reference theory of proper names on which their meaning just is their referent, he simply is not vulnerable to Wittgensteinian criticisms that proceed from those criticisms of the Augustinian picture of language. On the contrary, Raatikainen even suggests that Kripke's causal historical picture of names and natural kind terms can be incorporated with a 'meaning is use' outlook: "The practices of the linguistic community, which stretch back in time—the historical chains of communication and reference-borrowings—take care of reference, in Kripke's picture ... language users participate in a general

practice of reference-borrowing. One can perhaps even view it as a kind of “use” or “usage”, (...). ” (Raatikainen 2024: 14).

However, the idea that the moment of initial baptism fixes the name to its referent and that referent’s essential properties in all possible worlds, thus fixing the standards of correctness for all speakers’ uses of that name does run afoul of Wittgenstein’s criticisms of the Augustinian picture.

Part of the reason Raatikainen takes Kripke to escape the criticisms of the Augustinian picture to be really targeting to the following view: “a very strong notion cherished mainly by the logical positivists of the Vienna Circle in the 1930s (...) According to that view, ostensive definitions were supposed to be the ultimate source of all meaning and provide the foundation of both language and knowledge (...) every meaningful sentence was supposed to be reducible to elementary sentences, the latter in turn analyzable in terms of ostensive definitions of their primitive expressions, and in this way be all conclusively verifiable or falsifiable.” (Raatikainen 2024: 13).

It is true that this notion of ostensive definition falls within the scope of Wittgenstein’s intended target, but only because it is a very specific instance of the more general target. If Wittgenstein wanted to attack this specific, philosophically rich notion of ostensive definition he might have begun *PI* not with a quote from Augustine’s *Confessions*, but from, say, Carnap’s *Aufbau*. Wittgenstein, of course, did not do that, and with good reason. His target is more general than that very particular notion of ostensive definition, hence why he chose a much more familiar, and seemingly innocuous passage from Augustine. Therein Augustine is not claiming that *all* language is based upon “ostensive definitions (...) the ultimate source of all meaning (...) [to which] every meaningful sentence [is] reducible to elementary propositions, the latter in turn analyzable in terms of ostensive definitions of their primitive expressions, and in this way be all conclusively verifiable or falsifiable.” Augustine’s passage contains no such philosophical, verificationist bells and whistles, but is just a homely seeming remembrance of having things given in perception pointed to by adults, hearing the names of those things spoken and understanding that the word spoken named the object pointed to. But already, without any mention of verifiable elementary sentences or what have you, Wittgenstein’s target is fully in view.

It may have seemed innocuous, but Augustine in effect already described the child coming into their first language as though they already possessed a language, according to which their world given in their sense perceptions was carved up and categorized. All that remained was for the child to notice adults performing the familiar, antecedently understood gesture of pointing, and promptly learn that the spoken word labelled that indicated object in the public language they were learning. It was all along clear to the child, or at least very quickly became clear, what was being pointed out in every case – a particular object, a kind of object, a colour, a shape – and what the criteria for applying the relevant term in new cases were. We have here, rather than just a complaint about something so limited as protocol sentences in logical positivism, two of the main issues that will be treated throughout *PI* – the possibility of a private language and the rule following problem. Those are not just problems about “a very strong notion [of ostension] cherished mainly by the logical positivists”. Now, though that may establish that Wittgenstein had a wider target than Raatikainen took him to, it still doesn’t firmly establish the relevance of Wittgenstein’s attacks upon Kripke.

2. Grammar, Essence and the Necessary A Posteriori

This is where I believe some of the deepest disagreements between Kripke and Wittgenstein emerge. First though, let’s see how Raatikainen tries to harmonize the apparent differences here:

It is certainly true that in the austere framework of the early Wittgenstein’s TLP, there was definitely no place for the kind of separation of necessity and *a priori* and necessary truths which could only be known *a posteriori* that Kripke later suggested. With the later Wittgenstein, however, the situation might be a bit less straightforward. (Raatikainen 2024: 20)

It’s true that Wittgenstein’s views of necessity and the *a priori* shifted between *TLP* and *PI* (partly in that both notions come to admit of shifts). But various remarks from the later Wittgenstein on essence and essential properties suggest a sharp divergence from Kripke’s understanding of essence and necessity. Recall, on Kripke’s view that what we can discover in discovering a necessary *a posteriori* truth is a fact about the essence of a kind or individual,

where that essence is a *de re* essence – in the thing itself, not our descriptions of it.

Compare that though with how the later Wittgenstein discusses ‘essence’. In *RFM*, Wittgenstein states that “it is not the mark of a property that is ever ‘essential’, but rather the mark of a concept” (*RFM* I, §73). This is starkly opposed to Kripke’s *de re* understanding of essence, where an essential property is an essential property of an object itself, independently of how it is described or conceptualized. Something like this idea of Wittgenstein’s survives in *PI*, though there may be some subtle but important differences. I have in mind remarks such as: “*Essence* is expressed in grammar.” (*PI*, 2009, §371) And “...how can I decide what is an essential, and what an inessential, coincidental feature of the notation? Is there some reality lying behind the notation to which its grammar conforms?” (*PI*, 2009, §562)

Wittgenstein remarking that the case is similar to a proposal that some feature of the game of draughts is inessential goes on to say, “I am inclined to distinguish between essential and inessential rules in a game too. The game, one would like to say, has not only rules but also a *point*.” (*PI*, 2009, §564)

The suggestion I find here is that we should no more look to ground talk of essential features by looking to things themselves as opposed to investigating the point the notation serves in our lives in general than we do in the specific case of a game. Talk of an essential property in, say, chemistry does seem a good deal less arbitrary than in the case of game, but that worry is, I take it, supposed to be partially assuaged by remarks such as “The rules of grammar may be called “arbitrary”, if that is to mean that the *purpose* of grammar is nothing but that of language”. (*PI*, 2009, §497) Or, to return to *RFM*, “I say ... if you talk about essence you are merely noting a convention. But here, one would like to retort, there is no greater difference than that between a proposition about the depth of the essence and that about a mere convention. But what if I reply: to the *depth* that we see in the essence there corresponds the *deep* need for the convention.” (*RFM*, I, §74)

Since on the later Wittgenstein’s view, necessity and a priority are shiftable statuses, there is a way for accounting for the phenomena Kripke points to from within the Wittgensteinian scheme. In the later Wittgenstein, one of the

fundamental insights is the possibility of a flux between symptoms and criteria – the idea that something that is at one point an empirical concomitant of a phenomena or thing, can come to be the defining property of the thing. We do not however, discover that the latter property is the defining property of the thing, rather we come to use it as the defining property – either as the consequence of an explicit redefinition or a tacit shift in our use that we subsequently make explicit. In this way the synthetic effects what statements we consider definitional or analytic and a priori, and, also, which statements we will affirm in considering counterfactual scenarios. To fail to mark this distinction is, I believe, to fail to mark a distinction between a “shift in the water” and a “shift in the riverbed” that Wittgenstein speaks of in *On Certainty*:

96. It might be imagined that some propositions, of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and hard ones became fluid.

97. The mythology may change back into a state of flux, the river-bed of thoughts may shift. But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other.” (OC, 1975, §96, §97)

The movement of the waters I take it, corresponds to the fluid genuinely empirical propositions, whereas the river bed corresponds to the hardened a priori, necessary statements of the form of empirical propositions. The former are paradigm cases of a posteriori statements that are tested against experience and corroborated or not on that basis. They are conceivably false and not knowable a priori. The latter are a priori statements, not open to empirical testing and whose falsity is not conceivable to us, or is at least not a possibility we can presently make sense of. On Kripke’s view, a contingently a priori statement is gold is a yellow, metal has ultimately come to be falsifiable in just the same way as a false, straightforwardly empirical claim is. Likewise, a necessary a posteriori claim like gold is the element with atomic number 79 came to be confirmed in just the same way a contingently true proposition comes to be. No distinction is made here between “the movement of the waters and ... the shift of the bed itself”. Even though that distinction may not always

be sharp, I look at some consequences in the next section of erasing it altogether.

3. Criteria and the Problem of Transworld Identity in Kripke

One charge Raatikainen attempts to defend Kripke from is the accusation that Kripke failed to provide criteria for Transworld identification:

On Kripke's view, certain identities and other constancies across the scenarios can be built into the framework of counterfactual scenarios by stipulation: we reflect, for example, what could have happened to Nixon; that we are talking about Nixon is part of the framework, and the question of who is Nixon in a given scenario ("the problem of transworld identification") never arises. Or, if we want to consider how gold would behave in various counterfactual ... scenarios, we are talking about gold (the element with atomic number 79; give or take some tolerable degree of indeterminacy) and how it would behave—and not about a substance which merely looks and feels like gold: "Possible worlds" are stipulated, not discovered by powerful telescopes. (Kripke 1981: 44)

It is true that Kripke disclaims a Lewisian modal realism in which we need to provide criteria for identifying objects across different possible worlds. But Kripke cannot consistently maintain that we merely "stipulate" what is true in a possible world and consistently maintain that there are necessary a posteriori truths or de re essences and necessities.

Recall, on Kripke's view of natural kinds, gold always designated the element with atomic number 79. Even before we discovered this, prior to the discovery if we imagined a scenario where "gold" had some very different composition, we were merely imagining an epistemic possibility. We were conceiving of a metaphysically possible scenario in which we were confronted with an epistemic counterpart of some gold, and mistaking that for the possibility of gold not having its (then unknown) essential property. In this case, we hadn't yet stipulated any rule about what gold did or did not refer to in a possible world. Yet, the person of 1750, although they have stipulated a possible world in which gold is-such-and-such has not described a coherent possibility. So, possible worlds, and modal facts more broadly, on Kripke's view cannot be

simply be stipulated. One can describe a prima facie conceivable scenario, even one conceivable a priori, but fail to have described a genuine possibility. Even if one need not inspect possible worlds through a telescope in order to describe them, there are some modal facts to which our judgments of possibility and necessity are answerable. We do not get to simply stipulate possible worlds.

Furthermore, essences are properties of things themselves, and not things under a given description. If a thing has its essential properties in every possible world in which it exists, then I am not free to stipulate a possible world to the contrary – if I do so I speak falsely just as I would judging contrary to any other fact. Even if Kripke did not embrace modal realism and a counterpart theory of Transworld identification, he could not have consistently held that we stipulate possible worlds while also believing in de re essences and the necessary a posteriori. This tension in Kripke is, I believe, foreseen in the *Investigations*. First consider the many distinctions Wittgenstein draws therein between naming and describing, therein naming something is not yet to have described it, or said something true about it. That is merely a point about the grammar of naming. Kripke, however, takes this point and sublimates it in an attempt to explain it. The fact I can refer to someone by their name in a counterfactual scenario is explained by Kripke in virtue of the fact that names are rigid designators that pick out an object in all possible worlds where that thing exists and where it possesses all its essential properties. That is to try to turn naming into a special act of describing in which one tracks all or implicitly commits oneself to the existence of all and only that object's essential properties in the scenario described. But if one, with Wittgenstein, keeps naming distinct from describing and recognizes that what stipulations about counterfactuals we get away with doesn't depend upon our tracking essential properties of the objects themselves, but properties that simply allow one or one's audience to follow the counterfactual at hand, then one can avoid the worries about Transworld identity that Kripke wished to. But one can no longer maintain the picture of modality and essences he propounded.

References

Kripke, Saul (1981) *Naming and Necessity*. Oxford: Blackwell Publishing.

Raatikainen, Panu (2024) "On the Alleged Incompatibility Between Wittgenstein and Kripke", in: *Engaging Kripke with Wittgenstein: The Standard Meter, Contingent A Priori, and Beyond*, Martin Gustafsson, Oskari Kuusela, and Jakub Mácha, New York: Routledge, 9 – 27.

Wittgenstein, Ludwig (1975) *On Certainty*, G.E.M. Anscombe and G.H. von Wright (eds.), Denis Paul and G.E.M. Anscombe (trans.). Oxford: Blackwell Publishing.

Wittgenstein, Ludwig (1983) *Remarks on the Foundations of Mathematics*, Revised Edition, G.H. von Wright, Rush Rhes and G.E.M. Anscombe (eds.), G.E.M. Anscombe (trans.), Cambridge, MA: MIT Press.

Wittgenstein, Ludwig (2009) *Philosophical Investigations*. Revised 4th Edition, P.M.S. Hacker and Joachim Schulte (eds.), G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte (trans.), Oxford: Wiley-Blackwell.

Against the “Correspondence” Intuition

Susanna Melkonian-Altshuler (Vienna, Austria)

Abstract

Extended Abstract. Orthodoxy assumes that when we talk about the *dependence intuition*, we’re talking about an intuition that the correspondence theory is motivated by, so we’re just talking about the *correspondence intuition* (e.g., Horwich 1998, Hill 2002, Caputo 2013, Dodd 2018, Armour-Garb et al. 2023). The dependence intuition, going back to Aristotle’s *Categories* (14b, 9), is the intuition that one can rightly say that a statement is true because of how things (described by the statement) are in the world, but one cannot rightly say that these things in the world are a certain way because the statement is true. The content of this intuition is that the dependence of truth on reality is an *asymmetric explanatory relation*. Call the content *truth dependence*.

Truth dependence: $\langle p \rangle$ is true because p ; but it’s not the case that p because $\langle p \rangle$ is true.

While many use the two terms ‘dependence intuition’ and ‘correspondence intuition’ interchangeably, and many assume that the correspondence theory of truth can naturally account for the said intuition (e.g., Asay 2020), no one has provided a full-blown correspondence theorist account of truth dependence. Nor has anyone questioned the assumption that when we’re talking about the dependence intuition, we’re talking about an intuition that the correspondence theory of truth can naturally account for.

In this paper, I will question this claim, as correspondence is symmetric: x corresponds to y , and y corresponds to x . After illustrating that this is how early correspondence theorists, such as Russell (1912), and Moore (1910), understood the correspondence relation, I will move away from these traditional, say, pure correspondence theories of truth, to alleged impure correspondence theories of truths, e.g., truthmaker theory. Given that many take truthmaker theory to be a version of the correspondence theory of truth (e.g., Armstrong 1997) and take truthmaker theory to be asymmetric, I will investigate whether the correspondence theorist can make use of truthmaker theory to account for truth dependence. While giving the correspondence theory an excellent chance at succeeding, I will argue that it can’t. First, while truthmaking is asymmetric, that won’t help the correspondence theory to escape the symmetry of the correspondence relation. Correspondence remains symmetric even if we move to truthmaking.

In order to show this, I will begin by considering bridging principles offered by David (2009) that are supposed to bridge between symmetric correspondence and asymmetric truthmaking. I will argue that these are unsuccessful as they just merely posit that true propositions are not facts in order to get an artificial asymmetry off the ground via the principle of Non-Identity, which goes as follows:

necessarily, no proposition is identical with a fact; that is, for any x and y , if P_y & F_x , then $y \neq x$.

As a result, these bridging principles neglect fundamental questions that are at the core of the correspondence theory of truth such as ‘what are facts?’ Are they just true propositions?

After having established that David's bridging principles are useless with respect to getting a true asymmetry for the correspondence theory via truthmaking, I will move on to show that truthmaker theory as such is not a version of the correspondence theory of truth. While truthmaking comes in many varieties, e.g., object-based varieties, and fact-based varieties, even fact-based varieties of truthmaker theories are not varieties of the correspondence theory of truth. One reason for believing this is that truthmaking is an explanatory relation, but correspondence is not. It is a matching relation. To match one thing to another is not to explain one in terms of the other. For example, I can match a shoe to another shoe without explaining the one in terms of the other.

These are just some of the reasons for believing that truthmaker theory is not a path for the correspondence theory of truth to account for truth dependence. In the talk, I will present some more.

Does the Pauli Exclusion Principle Support Modalism or Essentialism?

Luke Miller (Vienna, Austria)

Abstract

In this paper I will evaluate the claims of modalism and essentialism in light of the Pauli exclusion principle (PEP). The PEP is the scientific principle that underlies the impenetrability of matter and its necessity has been exhaustively verified within quantum physics. Despite explaining a fundamental aspect of reality, its implications for current grounding debates have yet to be fully explored. The PEP governs a class of particles called fermions, which states that particles within the same system necessarily preclude each other from occupying the same state. I will argue that the PEP cannot be properly understood without reference to modality, vindicating modalism. However, I will also defend the position that the explanatory power of modal properties in quantum systems is asymmetric, and that this consideration favors the adoption of Non-Reductive Finean Essentialism. This view commits us to an irreducibly modal ontology while still affirming the importance of essences in metaphysical explanation.

1. Introduction

One of the most prominent debates within the philosophy of modality is between non-reductive modalism (hereafter simply modalism) and Finean Essentialism (FE). As the names suggest, modalism holds modality to be a basic and irreducible aspect of reality. This view was challenged in Kit Fine's seminal essay, "Essence and Modality", where he argued that modal properties must be grounded in non-modal essences (Fine 1994).

Since the publication of Fine's paper, debate has raged between essentialist and modalist interpretations of modality. Recently, modalists have developed a critique of FE, arguing that essence cannot adequately reduce modal notions. Briefly, any attempt to explain necessity in terms of an object's non-modal essential properties fails because they cannot explain why essential properties are necessary to begin with. As Mackie eloquently puts it: "It looks as if the account of essence in terms of real definition is intended to deliver a modal rabbit out of a non-modal hat. And I do not see how this can be done" (Mackie 2020).

In response to the critiques of FE discussed above, Wallner and Vaidya have proposed a reinterpretation of the Finean position: Non-Reductive Finean Essentialism (NRFE) (Wallner and Vaidya 2020). On this account, essences belong to the broader family of modal notions, playing an explanatory but not

ontologically reductive role. Because NRFE considers essences to have an irreducibly modal character, it is not susceptible to the arguments against the reductive version of FE described above (Wallner and Vaidya 2020: 9-10). I will argue in favor of this view by examining the Pauli exclusion principle (PEP), concluding that the PEP cannot be explained without reference to modality, but that modality must be non-reductively grounded in essence.

2. Justification for the Pauli exclusion principle

I have chosen to investigate the PEP for three reasons. First, there is a strong precedent within the philosophy of modality for deriving knowledge of necessity and possibility from scientific practice. Kripke emphasized the important role empirical observations can play in identifying the necessary properties of things picked out by rigid designators (Kripke 1980). Moreover, Bueno and Shalkowski extend the role of science in their defense of modalism, explicitly connecting our knowledge of modality with empirically determined probabilities (Bueno and Shalkowski 2015).

Second, The PEP has undergone stringent empirical verification, giving us good reason to believe that it describes a genuine necessity (Columbié-Leyva et al. 2023: 19).

Third, the PEP is minimally epistemic. By this I mean that it is largely insulated from debates over how to interpret quantum mechanics, allowing us to explore the modal implications of the PEP without being tied to an interpretation that may be later falsified. This is not true of other principles; for example, the Heisenberg uncertainty principle and wave-particle duality have spawned several interpretations, all in agreement with the data but postulating wildly different fundamental entities. Because I am interested in the metaphysics of modality, I have selected a principle whose interpretation is largely uncontroversial. Importantly, while the PEP is superior to other scientific principles in this regard, no scientific theory is completely free of epistemic notions. Relying on the concept of indistinguishability, the PEP is no exception. However, in quantum mechanics indistinguishability can be treated as a real property of objects, not a symptom of the imperfection of our observational capacity. I will relate how indistinguishability is a real constraint on particle behavior below.

3. What is the Pauli exclusion principle?

The Pauli Exclusion Principle states that for any set of indistinguishable fermions of the same kind, no two particles may occupy the same state simultaneously. It is the scientific principle responsible for what is colloquially understood as the impenetrability of matter. The PEP governs the kind of particles called fermions, and these include the fundamental particles that we associate with matter such as protons, neutrons, and electrons. Fermions are distinguished from another class of fundamental particles, bosons, which are not subject to the PEP. Consequently, an unlimited number of bosons may occupy the same state, making them ideal for carrying forces. A photon, or particle of light, is an example of a boson. All matter falls into these two classes, which are distinguished by the property of spin.

Spin, or more technically the internal angular momentum of a particle, has no classical counterpart. It is an intrinsic property of particles, the magnitude of which cannot change without changing particle identity. Like all quantum properties, a particle can only possess discrete spin values, which for spin come in half-integer units. Thus, particles may have either an integer or half-integer spin value. Fermions have half-integer spins; for instance, electrons have spins of $\pm 1/2$. Bosons have integer spins, such as a photon's spin of 0. This difference results in different symmetry properties, and therefore different behavior, of their respective wave functions.

Wave functions represent a particle's energy state and the square of the wave function can be interpreted as expressing the probability of finding the particle at a given location in space. When particles interact to form a system, their symmetries determine how their wave functions can combine. The specific type of symmetry that we are interested in is permutation symmetry, or the symmetry exhibited when two particles in a system switch places. A system of particles is described by the combination of all possible permutations of the system. In the simplest case describing a system of wave equation, Ψ , of two indistinguishable particles, x_1 , x_2 , this can be represented as:

$$\Psi(x_1, x_2) = (\psi_1(x_1)\psi_2(x_2) \pm \psi_1(x_2)\psi_2(x_1)) / \sqrt{2}$$

Whether the two expressions in the numerator are added or subtracted is determined by whether the particles are fermions or bosons. Bosons' wave functions are added, producing a symmetric wave function, while fermions' are subtracted, producing an anti-symmetric one (Landau 1974:157). From this equation it is evident that for fermions ψ_1 cannot equal ψ_2 if Ψ is to have a non-zero value. This is the mathematical basis for the PEP.

We can now also see why the PEP only applies to indistinguishable particles of the same kind. Within the wave equation, x refers to the particle as a pure quantum, the properties of that specific particle are determined by ψ . For this reason, x_1 cannot have a different value than x_2 without the particles belonging to different kinds. But what is to stop the particles being distinguished by their wave functions? This is in fact how we normally distinguish non-interacting particles. Only when particles interact to form a system do the particles become indistinguishable. Because the wave function of the system is the linear combination of all possible permutations, each particle, x_n , is described by each wave function $\psi_{1,2,\dots,n}$, so the basis for distinguishing the particles disappears.

4. The Modality of Exclusion

The representation above makes the PEP's necessity appear unremarkable; there are two classes of particles and two possible ways to combine wave functions, one of which leads to exclusion. However, the mathematical possibilities are only so restricted in one dimension. Once constraints on the number of dimensions and number of particles are removed, as occurs when electrons form atomic orbitals, so is the mathematical necessity of the PEP (Columbié-Leyva et al. 2023: 6-8). This has long been a source of consternation for physicists. As Feynman stated in a 1963:

Why is it that particles with half-integral spin are Fermi particles whose amplitudes add with the minus sign, whereas particles with integral spin are Bose particles whose amplitudes add with the positive sign? We apologize for the fact that we cannot give you an elementary explanation. An explanation has been worked out by Pauli from complicated arguments of quantum field theory and relativity. He has shown that the

two must necessarily go together, but we have not been able to find a way of reproducing his arguments on an elementary level. (Feynman et al. 1963; Columbié-Leyva et al. 2023: 5)

Feynman's consternation here stems from the appearance of an ungrounded physical necessity. Pauli proved that fermions cannot have symmetric wave functions, but his proof assumed that only symmetric and anti-symmetric wave functions were possible. This assumption is called the symmetrization postulate. Frustratingly, the symmetrization postulate explains what appears to be a brute physical necessity in terms of another brute physical necessity. Attempts to develop a purely mathematical justification for the connection between a particle's spin and the symmetry of its wave function have not yet been successful (Kaplan 2021: 4-5).

This development is exciting news for the modalist. In the PEP we have a fundamental principle in physics that is not explicable by the properties of the physical particles. Rather, the symmetrization postulate governs what conformations, and therefore properties, systems of particles can have. At least as far as our best science can tell us, we have found an ontologically irreducible natural modality.

5. Identity and the Symmetrization Postulate

While it does not appear to be the case the symmetrization postulate can be ontologically reduced, it may yet be the case that it can be non-reductively explained. Recent work done by Ilya Kaplan has developed an explanation for the symmetrization postulate based on particle identity. I will argue that if he is successful, then we have found an explanatory ground for the necessity of the PEP in particle identity, supporting NRFE claims about modality.

Because identity is such a philosophically loaded term, I will briefly examine Kaplan's use of the term before I discuss his proof. He understands identity as kind identity, with permutation symmetry as the sortal property determining kind membership. These kinds are defined by their permutation symmetry, which explains the energy properties of systems of identical particles. He justifies the introduction of kinds by proving that there can be no transition between states with different symmetries (Kaplan 2021: 10).

Once kinds are introduced, the necessity of the symmetrization postulate becomes evident. Suppose there is some system, S_n , of n particles that belong to kind, k , and k is neither the kind fermion nor the kind boson. Suppose now we add one particle of k to S_n . Kaplan shows that the permutation symmetry of S_{n+1} is the result of the linear combination of the wave function of S_n with that of the additional particle, but that attempts to do so violate particle indistinguishability. However, if we do select a wave equation in which the particles in S_{n+1} are indistinguishable, the permutation symmetry of S_{n+1} will not be the same as that of S_n . Because permutation symmetry determines kind identity, we would encounter a situation where adding a particle to a system changes the identity of all other particles in the system, and this is forbidden (Kaplan 2021:11-12). This proves that only bosons and fermions can preserve their kind identity within a system.

6. The PEP and Essence

Based on the argument above, I argue that we have good reason to affirm that the PEP is ultimately grounded in the identity of natural kinds as defined by their permutation symmetry. However, this appeal to natural kinds is neutral between modalism and essentialism, since our sortal properties merely identify criteria for kind membership apart from any modal claims. For the essentialist to be vindicated, we need the further supposition that the preservation of a system's permutation symmetry is grounded in the essence of what it is to be a particle of that kind. Interestingly, Kaplan believes that we are justified in making this stronger claim:

The system of particles with the definite permutation symmetry $[\lambda_N]$ must have some additional inherent particle characteristics that establishes why N -particles system *[sic]* is characterized by this permutation symmetry, like half-integer and integer values of particle spin for fermions and bosons. (Kaplan 2021: 10)

This appeal to the intrinsic property of spin favors an essentialist interpretation. The PEP is grounded in symmetry properties of fermions, which are grounded by the symmetrization postulate. This postulate is in turn

grounded in particle identity, which is grounded in a particle's essential property of spin. This essentialist interpretation of the PEP seems to be correct and favor the adoption of essentialism.

While the PEP can be grounded in the essence of fermions, treating this ground as ontological is unjustified. Let us examine the essential property of spin. As we noted above, spin is quantized and can only adopt values at half-integer intervals. But there is no reason grounded in the nature of spin that it could not have been quantized differently. Why weren't spins quantized as quarter-integers? If we answer that quarter-integer spin would permit permutation symmetries that do not preserve kind identity, then it turns out that spin itself is modal. The fact that spin is necessarily quantized in half-integers is just an aspect of the larger necessity that kind identity is preserved. This indicates that modalism is correct; modality is fundamental.

One might object that our inability to ontologically reduce the modality of the PEP to essence also indicates that we cannot explanatorily ground the PEP in essence. I do not believe this objection is warranted. While every property we appeal to will be irreducibly modal, it does not follow that they will all be equally explanatory. What it is for a particle to preserve its identity in a system is explained by having a specific property of spin and a specific symmetry of its wave function. Crucially, the explanation does not work in the other direction. What it is to have spin, even a specific magnitude of spin, is not explained by saying it preserves particle identity. That tells me what spin *does*, but it doesn't tell me what it *is*. This asymmetry indicates that essences do ground necessity in an explanatory sense.

7. Conclusion

We can now sketch a Non-Reductive Finean Essentialist account of the Pauli exclusion principle. The PEP states that any particle with half-integer spin will have an antisymmetric wave function. Such a particle belongs to the natural kind fermion defined by those two properties, and those properties constitute its essence. The necessity of the PEP is grounded in the necessity of kind identity preservation within system of particles, which restricts the possible spin values and symmetries a particle can have. The modalities described here are fundamental and irreducible. Despite this, the essences of fermions

explain what it is for its identity to be preserved: it is just the conservation of those essential properties without which it would not exist: half-integer spin and antisymmetry of its wave function. This investigation has shown that modalism and essentialism both capture important facets of reality and that the Pauli exclusion principle vindicates them both.

References

Bueno, Otavio; Shalkowski, Scott A. (2015) "Modalism and theoretical virtues: toward an epistemology of modality", *Philosophical Studies* 172, 671–689.

Columbié-Leyva, Ronald; López-Vivas, Alberto; Soullard, Jacques; Miranda, Ulises; Kaplan, Ilya G. (2023) "Symmetry of Identical Particles, Modern Achievements in the Pauli Exclusion Principle, in Superconductivity and in Some Other Phenomena", *Symmetry* 15(3), 701.

Feynman, Richard; Leighton, Robert; Sands, Matthew (1963) "Identical Particles", in: *The Feynman Lectures on Physics, Volume III*, New York: Basic.

Fine, Kit (1994) "Essence and Modality", The Second Philosophical Perspectives Lecture, *Philosophical Perspectives* 8, 1–16.

Landau, Lev; Lifshitz, Evgeny (1974) *Quantum Mechanics: A Shorter Course of Theoretical Physics*, Oxford: Pergamon.

Kaplan, Ilya G. (2021) "Modern State of the Pauli Exclusion Principle and the Problems of Its Theoretical Foundation", *Symmetry* 13(1), 21.

Kripke, Saul A. (1980) *Naming and Necessity* (Rev. Ed.), Oxford: Blackwell.

Mackie, Penelope (2020) "Can Metaphysical Modality Be Based on Essence?", in: Mircea Dumitru (ed.) *Metaphysics, Meaning, and Modality: Themes from Kit Fine*, Oxford: Oxford, 247–264.

Wallner, Michael; Vaidya, Anand J. (2020) "Essence, Explanation, and Modality", *Philosophy* 95(4), 419 - 445.

Acknowledgement

I would like to thank Dr. Anand Vaidya for his support and advice on this project. I am deeply indebted to his mentorship and insight.

LLM Collapse As Rule-Following Failure

Julia Minarik, Patrick Fraser (Toronto, Canada)

Abstract

Large Language Models (LLMs) like ChatGPT have an impressive command of human language: they can write poetry, ace the GRE (OpenAI 2023), and fool people into thinking they are human (Jones and Bergen 2023: 7). LLMs get good at using language by being trained on a high volume of human output. However, if trained on their own output, output that *seems human*, LLMs collapse, producing gibberish after a few iterations. We argue that this failure is not particular to LLMs. Rather, it arises from the essentially social character of language defended by Wittgenstein. When a member of a language community becomes isolated, unnoticed mistakes in their linguistic behaviour can teach them to speak in ways which violate the rules for proper use dictated by the practices of their community. We demonstrate that the Wittgensteinian community view of rule-following explains LLM collapse; moreover, LLM collapse provides support for the community view. We further note that LLM collapse reveals how the variability of human language ensures that it can continuously respond to our ever-changing reality.

1. Introduction

Large Language Models (LLMs) like ChatGPT gain an impressive command of human language by being trained on large quantities of human output: they can write poetry, ace the GRE (OpenAI 2023), and fool people into thinking they are human (Jones and Bergen 2023: 7). However, if trained repeatedly on their own output, output that *seems human*, LLMs produce gibberish. We argue that this failure (known as LLM Collapse) arises from the essentially social character of language. Per Wittgenstein, when a member of a language community becomes isolated, unnoticed mistakes in their linguistic behaviour can be amplified, causing the speaker to eventually violate the rules for proper use dictated by the practices of their community. We demonstrate that the Wittgensteinian community view of rule-following explains LLM collapse; moreover, LLM collapse provides support for the community view. Interestingly, viewing LLM collapse in this light reveals how the variability of human language ensures that it can continuously respond to our ever-changing reality.

2. LLMs and Collapse

LLMs generate textual outputs in response to prompts via next-token prediction: when supplied with a textual input, they predict the most likely

subsequent sequences of language particles called "tokens" (words, word parts, punctuation, etc.). LLMs predict tokens using statistical patterns of association they learn from the massive dataset of human-generated content they are pre-trained on (Brown et al. 2020: 8). They represent the individual tokens in this dataset as vectors in a high-dimensional vector space. Vectors which represent contextually related tokens like "king" and "queen" are closer than unrelated ones; semantic relations are encoded geometrically. Since these models are pre-trained, after training, the vector space representation of the model is fixed. If this representation needs to be adjusted, the model must be trained again. Notably, while LLMs model human natural language use, they do not do so by "hard coding" the presumed structure of language as studied by linguists into their models. Next-token prediction does not build in any of the "common normative goals of language use" (Millière and Buckner 2024: 8) such as our cooperative aim to state truths, or obey conversational maxims.

It has recently been shown that when LLMs are trained on their own output, rather than human-generated output, they exhibit "collapse." The LLM used to demonstrate collapse is OPT-125M, the predecessor of OPT-175B, the open-source model that performs most closely to GPT-3 (Shumailov et al. 2023: 3; Zhang et al. 2023: 2). After only nine iterations of retraining on its own output, OPT-125M fails to use language in a way that is meaningful to human interpreters, see Figure 1.

Example of text outputs of an OPT-125m model affected by *Model Collapse*— models degrade over generations, where each new generation is trained on data produced by the previous generation.

Input: some started before 1360 — was typically accomplished by a master mason and a small team of itinerant masons, supplemented by local parish labourers, according to Poyntz Wright. But other authors reject this model, suggesting instead that leading architects designed the parish church towers based on early examples of Perpendicular

Outputs:

Gen 0: Revival architecture such as St. John's Cathedral in London. The earliest surviving example of Perpendicular Revival architecture is found in the 18th @-@ century Church of Our Lady of Guernsey, which dates from the late 19th century. There are two types of perpendicular churches : those

Gen 1: architecture such as St. Peter's Basilica in Rome or St. Peter's Basilica in Buenos Aires. There is no evidence that any of these buildings were built during the reign of Pope Innocent III, but it is possible that they may have been built during the reign of his successor, Pope Innocent

Gen 7: architecture in England. In an interview with The New York Times, Wright said : " I don 't think there is anything wrong with me being able to do what I want to do. It just doesn 't work for me. " He added : " I don 't know if you can call it funny,

Gen 9: architecture. In addition to being home to some of the world's largest populations of black @-@ tailed jackrabbits, white @-@ tailed jackrabbits, blue @-@ tailed jackrabbits, red @-@ tailed jackrabbits, yellow @-

Figure 1: *Model Collapse using OPT-125M, image taken from Shumailov et al. (2023, pg.3)*

Collapse has been described as "a degenerative process...where generated data end up polluting the training set of the next generation of models; being trained on polluted data, they then mis-perceive reality" which leads to models "misinterpreting what they believe to be real, by reinforcing their own beliefs" (Shumailov 2023: 3). Setting aside the belief talk for the sake of theoretical neutrality, the suggestion that these models begin to misperceive (or misrepresent) reality by reinforcing their own behaviour is something we consider below. More technically, model collapse is described as follows:

(...) over time we start losing information about the true distribution [that codifies correlations in human language use], which first starts with tails disappearing, and over the generations learned behaviours start converging to a point estimate with very small variance. (Shumailov 2023: 2)

LLMs are built to predict subsequent linguistic tokens. At each point during sentence generation, they examine a distribution of possible next tokens and choose one of the most likely tokens. Over time, the tokens with the lowest conditional probabilities, those that capture idiosyncratic or "creative" uses of language (jokes, novel metaphors, new words and concepts, and mistakes) get generated less and less often. Then, when the model is retrained on its own

output (in which these unlikely outputs probably do not appear), it updates its statistical distribution over the new dataset and the probabilities associated with next tokens are readjusted. Eventually, token sequences with the lowest probabilities drop out entirely. This is described as "probable [utterance] events being overestimated, and improbable events being underestimated" (Shumailov 2023: 2). When this process is iterated, the support of the distribution from which probable next tokens are predicted converges to a point, making only a small number of subsequent tokens probable. The takeaway is that when retrained on their own output, the LLMs are changing the rules for how they generate language in a way that indiscriminately amplifies patterns in their past behaviour.

3. Collapse as Rule-Following Failure

We now show that LLM collapse is not merely a defect of LLMs but, rather, an instance of a more general phenomenon. Specifically, we contend that Wittgenstein's rule-following considerations explain why LLM collapse occurs: it is an instance of a general phenomenon that is expected of any isolated language user whatsoever (flesh-bound, digital, or otherwise).

LLMs are good language users. They successfully produce linguistic responses that conform to the rules of human natural language. (We remain neutral on whether LLMs have minds, consciousness, or understanding, and require only that they produce coherent, meaningful, contextually appropriate, and frequently true natural language responses.) What we aim to explain is why this ability deteriorates as they are iteratively trained on their own outputs.

To explain this, we follow Wittgenstein in adopting the general view that the meaning of language is determined by rules which govern its proper use (BBB 1998; PI 2009: §43). We additionally adopt the more specific Wittgensteinian view that these rules are essentially social; they are the customs of a linguistic community. Specifically, the proper uses of a particular element of \mathcal{L} are those which the \mathcal{L} -community would customarily allow. As Wittgenstein writes, "a person goes by a signpost [i.e. follows a rule] only in so far as there is an established usage, a custom" (PI 2009: §198) and "[t]o follow a rule, to make a report, to give an order, to play a game of chess, are customs (routines, institutions). To understand a sentence means to understand a language. To

understand a language means to have mastered a technique" (PI 2009: §199); see also (PI 2009: §202).

We must disentangle two issues, namely, rule-following and private language. Wittgenstein's rule-following skepticism motivates the view that meaning is socially determined. One may separately ask whether a speaker can meaningfully speak a language in private (i.e. without ever interacting with a broader community). Some argue that meaningful private language is impossible (Kripke 1982; Malcolm 1986; Malcolm 1989; Savigny 1991); others disagree (Baker & Hacker 1984; Baker & Hacker 1985), see (Boghossian 1989; Canfield 1996) for details of this debate. We are only concerned with what happens when a member of an established language community becomes isolated from that community, and the extent to which they will continue speaking that language. Thus, the speakers we are interested in are not putative private language speakers and so we are neutral towards the possibility of private language.

Speakers of a language \mathcal{L} learn how to speak following the \mathcal{L} -rules by producing linguistic behaviour and being met with responses from other \mathcal{L} -speakers, approving or disapproving of their behaviour (Kripke 1982: 89-100; PI 2009: §6-9). Whether reflecting on these responses or not, the learner's behaviour is regulated by the approvals and disapprovals of the \mathcal{L} -community, and learning a language amounts to learning the customs which fix the rules governing that language. If a speaker's linguistic behaviour deviates too far from the customs of the \mathcal{L} -community, they cannot be called an \mathcal{L} -speaker.

We now consider what happens when a member of a particular language community becomes isolated from that community. Concretely, we say that an \mathcal{L} -speaker becomes linguistically isolated if, already being a member of the \mathcal{L} -community, they become disconnected from \mathcal{L} -community in a manner which renders bidirectional communication between the speaker and the community impossible. Linguistic isolation can have many guises: A speaker who is stranded on a desert island, travelling in a foreign country, or talking in their sleep beside their spouse may all be isolated. All that is required is some obstruction preventing \mathcal{L} -community from regulating the speaker's language use whence the perlocutionary force of the \mathcal{L} -community's endorsements is suppressed.

If an isolated speaker is to continue to follow the \mathcal{L} -rules, they must regulate their behaviour themselves. How might one do this? In the absence of external monitoring, the only way for a speaker to ensure that their linguistic behaviour is rule-following is to reflect and self-correct when they notice mistakes. However, speakers rarely reflect on their own linguistic behaviour; in general, they speak without hesitation or deliberation (Kripke 1982: 87; PI 2009: §212, §219). We rely heavily on others to regulate our linguistic behaviour and we typically assume that our behaviour is rule-following in the absence of disapproval.

Yet speakers do, of course, make mistakes. Indeed, empirical studies suggest that "other-initiated repair signals" whereby a human speaker's audience requests a correction to an utterance for communication to be successful is extremely frequent—the average across 12 languages is approximately once every 1.4 minutes (Dingemanse et al. 2015: 4). Human language users constantly monitor the language use of their interlocutors, and speakers rely on this monitoring, but this monitoring is unavailable for isolated speakers. If speakers catch themselves in error, they may correct themselves (e.g. if someone is counting something, loses track, and starts over). However, we typically delegate the regulation of our linguistic behaviour to others; reflective self-correction is usually passive. Hence, we should expect isolated speakers to produce linguistic behaviour that fails to obey the customs of their language community without these mistakes being subjected to the usual disapprovals (or other-initiated repair signals) that would be expected were the speaker not isolated.

This shows that we cannot expect an isolated \mathcal{L} -speaker to successfully evaluate their own \mathcal{L} -use as rule-following; they will permit their own uncorrected language misuse. On its own, this is not that surprising. Language users make mistakes. However, a speaker's language use is regulated by and dynamically evolves following whatever community monitoring they are exposed to. In permitting their own mistakes to go uncorrected, an isolated speaker will learn to display linguistic behaviour that increasingly diverges from the rules enforced by the community they originate from. Rhees writes that "[i]t seems that in a private language everything would have to be at once a statement and a definition" (Ayer & Rhees 1954: 83). This seems true also for the isolated speaker.

If private language is impossible, the isolated \mathcal{L} -speaker's linguistic behaviour devolves into meaninglessness. If private language is possible, the isolated \mathcal{L} -speaker could be found to passively construct and use a new language distinct from \mathcal{L} of which they are the only speaker without noticing. In either case, the isolated speaker stops speaking \mathcal{L} .

The shift undergone by the isolated speaker need not make their future linguistic behaviour unrecognizable to the \mathcal{L} -community. They may keep much of the same vocabulary, and they may continue to follow some \mathcal{L} -rules and not others. Our suggestion is rather that the language they speak after sufficiently long isolation will likely no longer coincide with the language spoken by their community of origin. Our general conclusion about language is this: If a member of the \mathcal{L} -community, becomes linguistically isolated from the \mathcal{L} -community, we should expect that, over time, their \mathcal{L} -use will depart from what the \mathcal{L} -community would endorse, were they present.

While this phenomenon is to be expected of any user of any language, we claim in particular that LLM collapse is an instance of this phenomenon. Insofar as LLMs are language users, they are members of the language communities associated with the human languages they use. Humans are generally prepared to endorse LLM outputs as meaningful and approve of their behaviour as rule-following. This should be unsurprising since LLMs are trained on a large corpus of human-approved instances of human language use.

However, one difference is clear: human linguistic behaviour is constantly monitored by other members of the linguistic community and regulated in real time through the dynamic approval and disapproval of other people. For LLMs, however, the regulation of their linguistic behaviour occurs only intermittently, namely, when they are retrained. After an LLM is deployed, its learning ends and its linguistic behaviour becomes unresponsive to the approvals or disapprovals of other language users (aside from reinforcement which plays a role in improving LLM conversational pragmatics, but not semantics). In this way, once an LLM is deployed for use, it becomes linguistically isolated from its language community until it is retrained next.

Despite their isolation, LLMs avoid the pitfalls described above under ordinary operating conditions because even if they make undetected mistakes in their language use, they do not amplify these mistakes. They lack a mechanism for dynamically revising how they produce their linguistic behaviour, and when they are retrained on human-output these mistakes are "corrected" by human language use data. If, however, an LLM is trained on its own outputs, this may be interpreted as the LLM approving of its own linguistic behaviour—whether or not the human language community would agree—and amplifying these approvals so as to learn to produce similar behaviour in the future, learning to mimic itself.

Since LLMs are linguistically isolated, the possibility of mistakes becoming amplified through the unreflective self-approval of their linguistic behaviour suggests that when they are iteratively trained on their own outputs, they will degenerate into producing outputs that fail to conform to the rules of use for human language. This, however, is exactly the semantic failure displayed in LLM collapse.

4. Language and Reality: Ours and Theirs

Human language is for communicating in and about the world, and so its rules are constantly evolving in response to changes in our world—they are externally repaired by it. We are physical beings that interact with an external environment which we do not entirely know, which changes on its own, and which we change through our actions. As the circumstances facing a linguistic community change, so will their collective beliefs, desires, expectations, and intentions. Our language must evolve with our scientific understanding of the world, and accommodate our creative acts which alter our reality artistically, technologically, and socially. If the rules governing language use were fixed absolutely, as they are for a pre-trained LLM, the resulting language would be too rigid to accommodate such shifting goals. Injecting creativity into possible uses of language—creating new hermeneutical resources for expression—serves a practical need.

As both Wittgenstein (PI 2009: §80) and Waismann (McKinnon et al. 1945: 121-122) argue, many (perhaps all) of the concepts language allows us to express are necessarily underdetermined. Specifically, since there are always

possible ways things could be that cannot be anticipated when language rules are adopted, there exist scenarios in which it is indeterminate whether or not our concepts apply. Importantly, even concepts whose rules of applicability are not vague can exhibit this sort of "open texture," for the domain over which those rules must be specified cannot be anticipated in advance. When faced with new putative applications of a concept, new customs must be established.

How we conceive of our world shapes the customs we establish and determines the language we have available, but these customs are always renegotiable. For LLMs, however, the limits of *our* language are the limits of *their* world. LLMs do not have direct contact with our reality. They are trained on our language and then become fixed; unless they are retrained they cannot re-engineer their concepts. Although LLMs map relations between language particles that track physical features of our world, they are unresponsive to changes in that world—they are bound by their statistical distribution over language token associations, and if they are to incorporate new terms, uses and concepts, they must be retrained. This is why, as Shumailov and colleagues suggest, LLMs "mis-perceive" reality during collapse: the pictures of reality they get from their own output are progressively less variable than reality itself.

Interestingly, as LLMs and their outputs begin to populate the internet, they will begin to produce part of the changing world that we need to respond to. As Brinkmann et al. highlight, "[h]umans and machines contribute to this existing cultural variation...However, machines, leveraging their unique capacities, can produce traits distinct from those produced by humans, thus potentially steering culture towards new paths" (2023: 1859). LLMs produce output that is in our world and they do this quickly. We must be careful not to narrow our own language use in response to their lack of variability. It is in this variability that human creativity thrives. Renze and Guzen have called the more unconventional and less likely linguistic outputs of LLMs the "creative ones" (2024: 2) and creativity is often said to require either novel or surprising outputs (Gonzalez & Haselager 2005; Boden 2009; Stokes 2011; Nanay 2014). Think of the beautiful and individual turns of phrase in literary contexts. These unlikely uses of language are among those that drop out first during collapse. Without community monitoring, variability disappears. Exploring the

relationship between community monitoring, linguistic variability, and creativity in LLMs would be an exciting way to move forward.

References

Ayer, A. J.; R. Rhees (1954) "Symposium: Can There Be a Private Language?", *Proceedings of the Aristotelian Society, Supplementary Volumes* 28, 77–94.

Baker, Gordon P.; P. M. S. Hacker (1984) *Investigations: Scepticism, Rules and Language*, Blackwell.

Baker, Gordon P.; P. M. S. Hacker (1985) *Wittgenstein: Rules, Grammar, and Necessity*, Blackwell.

Boden, Margaret A (2009) "Computer Models of Creativity", *AI Magazine* 30 (3), 23–34.

Boghossian, Paul A (1989) "The Rule-Following Considerations", *Mind* 98 (392), 507–549.

Brinkmann, Levin; Baumann, Fabian; Bonnefon, Jean-François; Derex, Maxime; Müller, Thomas F.; Nussberger, Anne-Marie; Czaplicka, Agnieszka; et al. (2023) "Machine culture", *Nature Human Behaviour* 7 (11), 1855–1868.

Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; et al. (2020) "Language Models are Few-Shot Learners", arXiv: 2005.14165 [cs.CL].
Canfield, John V (1996) "The Community View", *The Philosophical Review* 105 (4), 469–488.

Dingemanse, Mark; Roberts, Seán G; Baranova, Julija; Blythe, Joe; Drew, Paul; Floyd, Simeon; Gisladottir, Rosa S.; et al. (2015) "Universal Principles in the Repair of Communication Problems", *PloS One*.

Gonzalez, Maria Eunice Quilici; Haselager, Willem (Pim) Ferdinand Gerardus (2005) "Creativity: Surprise and Abductive Reasoning", *Semiotica* 2005 (153), 325–342.

Jones, Cameron; Bergen, Benjamin (2023) "Does GPT-4 Pass the Turing Test?", arXiv: 2310.20216[cs.AI].

Kripke, Saul A. (1982) *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Harvard: Harvard University Press.

Mackinnon, D. M.; Waismann, F.; Kneale, W. C. (1945) "Symposium: Verifiability", *Proceedings of the Aristotelian Society, Supplementary Volumes* 19, 119–150.

Malcolm, Norman (1986) *Nothing is Hidden: Wittgenstein's Criticism of His Early Thought*, New York, NY, USA: Blackwell.

Malcolm, Norman (1989) "Wittgenstein on Language and Rules", *Philosophy* 64 (247), 5–28.

Millière, Raphaël; Buckner, Cameron (2024) "A Philosophical Introduction to Language Models– Part I: Continuity With Classic Debates", arXiv:2401.03910 [cs.CL].

Nanay, Bence (2014) "An Experiential Account of Creativity." in: Elliot Samuel Paul and Scott Barry Kaufman (eds.) *Philosophy of Creativity: New Essays*, 17–36.

OpenAI; Achiam, Josh; Adler, Steven; Agarwal, Sandhini; Ahmad, Lama; Akkaya, Ilge; Aleman, Florencia Leoni, et al. (2023) "GPT-4 Technical Report", arXiv:2303.08774 [cs.CL].

Renze, Matthew; Guven, Erhan (2024) "The Effect of Sampling Temperature on Problem Solving in Large Language Models", arXiv:2402.05201v1 [cs.CL].

Savigny, Eike V (1991) "Self-Conscious Individual versus Social Soul: The Rationale of Wittgenstein's Discussion of Rule Following", *Philosophy and Phenomenological Research* 51 (1), 67–84.

Shumailov, Ilia; Shumaylov, Zakhar; Zhao, Yiren; Gal, Yarin; Papernot, Nicolas; Anderson, Ross (2023) "The Curse of Recursion: Training on Generated Data Makes Models Forget", arXiv:2305.17493 [cs.LG].

Stokes, Dustin (2011) "Minimally Creative Thought", *Metaphilosophy* 42 (5), 658–681.

Wittgenstein, Ludwig (2009) *Philosophical Investigations*. 4th ed. P. M. S. Hacker and Joachim Schulte (eds.). G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte (trans.).

Wittgenstein, Ludwig (1991) *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigation'*. 1st edition. Malden, Massachusetts: Wiley-Blackwell.

Zhang, Susan; Roller, Stephen; Goyal, Naman; Artetxe, Mikel; Chen, Moya; Chen, Shuohui; Dewan, Christopher, et al. (2022). "OPT: Open Pre-trained Transformer Language Models", arXiv:2205.01068 [cs.CL].

Acknowledgement

The authors wish to thank Jennifer Nagel and Cheryl Misak for their valuable comments. PF and JM are both supported in part by funding from the Social Sciences and Humanities Research Council.

Coreference and Cofundamentality

Yang Minghui (Amherst, USA)

Abstract

Take two terms that have the same semantic value. For a concrete example, we may use ‘earlier than’ and ‘later than’: they are converses of each other, and thus arguably they are coreferential. Could the notion of fundamentality distinguish them? For example, could ‘earlier than’ turn out to be fundamental, but not ‘later than’? I argue that there is no reason to believe in such distinctions. Coreferential terms are co-fundamental if any of them is fundamental at all. In section 1 of the paper I clarify the notion of fundamentality, and in section 2 and 3 I address two possible reasons that may lead us to distinguish between ‘earlier than’ and ‘later than’: minimality and perspicuity. I argue that neither option is effective.

Take two terms that have the same semantic value. For a concrete example, we may use ‘earlier than’ and ‘later than’: they are converses of each other, and thus they arguably refer to one and the same relation (Fine 2000). Could the notion of fundamentality distinguish them? For example, could ‘earlier than’ turn out to be fundamental, but not ‘later than’?

I venture to say that the answer is ‘no’: they are both fundamental. More generally, coreferential terms are all fundamental, if any one of them is fundamental at all. I will call this the thesis of co-fundamentality. In what follows I will clarify the notion of fundamentality (section 1), and then defend the thesis of co-fundamentality by arguing against some reasons to think otherwise (section 2-3).

1. Fundamentality

What exactly does it mean to say that a term is fundamental? There are broad-brushed slogans: fundamental terms “carve nature at its joints”, “capture the structure of reality”, etc. There are also relatively clear examples: ‘quark’ is fundamental, ‘grue’ on the other hand is not fundamental but gerrymandered. But when we try to pin these suggestions down in precise terms, a variety of views will arise, and they can be categorized in various ways.

McSweeney (2019: 120) gives us one categorization. She divides between ontic vs. ideological conceptions of fundamentality. The ontic conception of fundamentality defines the fundamentality of a term as (reducible to) the

fundamentality of its semantic value. To use the old example, we might say that ‘quark’ is fundamental because the property of being a quark is fundamental, while ‘grue’ is not fundamental because the property of being grue is not fundamental. Exactly what makes for a fundamental property is up for grabs: we may supplement the theory by Lewis’s theory of naturalness (1983), or a theory of universals, but those details need not concern us.

The ideological conception on the other hand does not seek to reduce everything to the fundamentality of some entity. Rather, it claims that ideological commitments differ in kind from ontological commitments. For an example of this Sider (2011: 92) creates the fundamentality operator F , and says the following with respect to its meaning:

To say $F(\text{and})$ is not to say something about an alleged object Conjunction. It is not to say anything about any thing at all. It is nevertheless to say something true, something objective, something about reality. (Sider 2011: 92)

The suggestion is that by applying ‘ F ’ to ‘and’ we are engaged with a *suis generis* fact that goes beyond the object-property distinction. We are not attributing any property to any entity—not the semantic value of ‘and’, nor the word ‘and’ itself, *qua* linguistic item. (The second point is made explicit by Sider 2011: 91. We may also see it by attending to the fact that the word ‘and’ is used rather than mentioned in the expression ‘ $F(\text{and})$ ’.)

The ontic vs. ideological distinction, however, is just one way to categorize the relevant views. I suspect that it is not the best way for my purposes. Why?

We first observe that the ontic conception makes the co-fundamentality thesis true. For the identity of the semantic value is the only thing that matters to the fundamentality of a term, and coreferential terms have identical semantic value. On the flip side, any difference concerning fundamentality must be accompanied with semantic differences. To use an example from Sider (2020: 182), suppose someone says that mass-quantities measured by kilograms are more fundamental than by pounds, this position is coherent under the ontic conception only if different measurements of mass invoke different entities:

for example, we might say that metric measurements invoke a function from objects to numbers that is different from the function used by imperial measurements. So far, so good.

But things are more complicated under the ideological conception. While the direct inference from the sameness of the semantic value to co-fundamentality is not licensed under the ideological conception, the co-fundamentality thesis may nonetheless be true. Semantics may go hand in hand with fundamentality.

Take nominalism about properties. It is incompatible with the ontic conception of fundamentality because there are no semantic values to play with. However, it is compatible with the idea that the metaphysics of fundamentality is somehow constrained by semantic facts. Define the semantic contribution of a predicate by how it is paired with referring terms to express facts: when ‘quark’ is paired with the name ‘Quarkie’ it expresses the truth ‘Quarkie is a quark’, but it does not express any fact when it is paired with ‘Elie’ (because Elie is an electron), and so on and so forth. Now, the nominalist may claim that terms that have the same semantic contribution do not differ with respect to their fundamentality; on the flip side, any difference in fundamentality with respect to some ideology must be accompanied by semantic differences. This won’t follow from the ideological conception of fundamentality, but it is at any rate a coherent position. Indeed, such a nominalist may easily argue that ‘earlier than’ and ‘later than’ are co-fundamental: intuitively, the fact that 2023 is earlier than 2024 and the fact that 2024 is later than 2023 are one and the same fact (cf. Fine 2000: 3-4 who utilizes exactly this intuition to argue for co-reference), and by generalizing from this pattern we will see that ‘earlier than’ and ‘later than’ are involved in the representation of the same range of facts. The nominalist will conclude that ‘earlier than’ and ‘later than’ have the same semantic contribution. They are co-fundamental if either of them is.

Of course, under the ideological conception it is also possible to make the co-fundamentality thesis false. We may do this by denying the alleged connection between semantics and fundamentality. The thought could be that just contributing to the representation of a fundamental fact is not enough, the term must contribute in the right way. This more demanding notion of fundamentality leaves room for saying that ‘earlier than’ is a bit of the fundamental ideology but ‘later than’ is not, perhaps because only ‘earlier

than' represents the structure inherent to the relevant fundamental facts in the right way. Exactly what this means is an open question, and I will come back to it later.

Thus, we may divide the ideological conception into a semantically-constrained version and an extra-semantic version. However, it is easy to see that the semantically-constrained version of the ideological conception is continuous with the ontic conception in spirit, if not in appearance. Thus we may expand the word "semantically-constrained" to cover the ontic conception as well. This is why I said earlier that I suspect the ontic vs. ideological division is not the best categorization. The better division is semantic vs. extra-semantic. The semantically-constrained view claims that the notion of fundamentality is no more fine-grained than the relevant semantic notions. Extra-semantic view on the other hand allows metaphysics to draw distinctions at places where semantics does not. These views could be distinguished by how they would deal with the case of coreferential terms.

But which view is correct—the semantic, or the extra semantic? Do we really have reasons to believe that co-referential (or semantically-coordinated, if we go with the aforementioned nominalism) terms could differ with respect to their fundamentality? I submit that there is no significant reason. I now move on to disarm some reasons one might use to support the extra-semantic difference. Ultimately, this will give us an argument that the extra-semantic view is not well supported.

2. Minimality

One prominent argument against the idea that 'earlier than' and 'later than' are both fundamental appeals to the idea of minimality. The thought is that the collection of fundamental properties (etc) should be just enough for generating a complete description of what is going on in reality (Lewis 1986: 60). Anything more than this is redundant. Now, it seems that just having one of 'earlier than' or 'later than' is enough for a complete (B-)theory of time. So, having them both in our fundamental ideology is redundant.

I do not find this argument convincing. While it could be granted that redundancy is a bad thing, having both 'earlier than' and 'later than' in the

fundamental ideology is not redundant given their co-reference. In what follows, I expand on this response by addressing two possible reasons to think that redundancy is bad and minimality is good: that it goes against our modal intuitions, or that it is not parsimonious.

Take the modal argument for minimality: redundancy is bad and minimality is good because, if there were redundancy on the fundamental level, then there would be unexplained necessary connections on the fundamental level. Some versions of this argument directly utilize semantic notions. As an example, McSweeney (2019: 125) uses the following principle:

Weak non-redundancy: there are no unexplained necessary connections between fundamental facts.

This principle has no force in dealing with the case of coreferential terms. For coreferential terms contribute to the representation of the same facts, and there won't be any problem about a fact being necessarily connected to itself! Thus, the co-fundamentality of 'earlier than' and 'later than' will not cause any violation of McSweeney's principle.

One might try to avoid this problem by talking about fundamental truths instead of facts:

Weak non-redundancy*: there are no unexplained necessary connections between fundamental truths.

But the improvement is superficial. For in the case of both 'earlier than' and 'later than' being fundamental, there is a ready-made explanation of why there are fundamental truths in necessary connection: the truths are connected because they are semantically coordinated: they represent the same fact, the same aspect of reality. Thus, once again we do not have any violation of the principle, and for that matter there is no argument against the co-fundamentality of 'earlier than' and 'later than' either.

It should be noted that my response does not challenge the modal intuitions themselves. That will give us another response: we might say that the relevant modal intuitions should not be taken too seriously (see Sider 2011: 218-9). I am sympathetic to that response, but I do not commit myself to it. For I take my response to be stronger: even if we treat the modal intuitions seriously, the

argument against the co-fundamentality of coreferential terms is nonetheless ineffective. For those modal intuitions bear on semantic notions, and thus they do not allow for distinctions more fine-grained than semantics, as the extra-semantic view on fundamentality would require.

So much for the modal argument. Let us now move on to the second argument for minimality, which appeals to the notion of parsimony. The thought is roughly that a theory with a smaller number of primitives is more parsimonious than a theory with a bigger number of primitives. Thus, minimality is good, and redundancy is bad. My response to this argument is parallel to my previous response to the modal argument: I do not find the theory that has both ‘earlier than’ and ‘later than’ objectionably unparsimonious, given that the terms are coreferential.

Indeed, it is not clear to me that parsimony will allow us to draw distinctions more fine-grained than our semantics would allow. The background theory is lacking: why must we measure parsimony by counting the number of individual terms rather than their semantic equivalent classes? By adding a term that is synonymous with some pre-existing term, we do make the list of vocabularies a little longer, but we do not complicate any of the facts being represented. It is not very plausible, then, to say that a smaller number of words must represent reality better than a bigger number of words. We may instead adopt a semantic-based view of parsimony under which theories that are semantically equivalent become equally parsimonious.

(One view in the vicinity of my suggestion is Sam Cowling’s view that parsimony is measured by counting the types of ideological commitments made by a theory. For we may consider the narrowest notion of a type to be a semantic type under which coreferential terms are put together. I will skip more details of Cowling’s view. See Cowling 2013 for more details.)

I submit that there is no good argument against the co-fundamentality of ‘earlier than’ and ‘later than’ from the constraint of minimality. As a final caveat, I want to note that my response could be extended to disarm the notion of arbitrariness.

It is usually suggested that arbitrariness works in the opposite direction of redundancy: while having too many fundamental things is redundant,

dropping some of them could be arbitrary, which is also bad. Thus redundancy and arbitrariness become two vices that trade off with each other. Now, I have argued that having a pair of coreferential terms is not redundant in any objectionable sense because adding coreferential terms does not complicate what is being represented by the relevant theory. But the opposite claim is also true: by dropping one of the coreferential terms we won't do any damage to the representation either. The redundancy-arbitrariness tradeoff simply does not matter in the case of the relevant terms being co-fundamental.

3. Perspicuity

An alternative argument against the thesis that coreferential terms are co-fundamental appeals to the idea of perspicuity. The idea is that while the relevant terms might latch onto the same entities/facts their representation might fail to be equally perspicuous.

Here is the intuitive idea of perspicuity. Think of picturesque representations of what is going on in reality. Say, that there is a natural landscape, and there is a painting of it. There is, then, a fact of the matter concerning how similar the painting is to the landscape. This is to be determined by how well the painting recovers the internal structures in the natural landscape: if a house is located on the top of a hill, the painting should contain a painted house on the top of the painted hill, etc. We may extend this lesson to the case of representations in general. For example (see McSweeney 2019: 128-9 for this discussion), if the structure of a fact is [P AND Q], then to represent that fact by 'P and Q' is better than representing it as 'It is not the case that not-P or not-Q' because the former captures the conjunctive structure in the fact being represented, but not the latter. The same logic will apply to the case of 'earlier than' and 'later than': they latch on the same facts, but one might argue that those facts might have internal structures. If that structure favors, say, 'earlier than' over 'later than', that would explain why 'earlier than' is fundamental but 'later than' is not.

In order to fully develop this view we must answer the following two questions:

Question 1. What is the feature of the relevant temporal facts that would help us distinguish ‘earlier than’ from ‘later than’?

Question 2. By virtue of what might we say that the feature is captured, or reflected, by ‘earlier than’ but not ‘later than’?

I suspect these answers are not independent but should be given all at once. But, at any rate, these are intractable questions. It is very hard to imagine what the answers will look like. But could we at least come up with some reasons to believe that some answer will be available?

I want to say I am skeptical. Here are some reasons for my skepticism.

First of all, perspicuity talk should not be another way of stating the extra-semantic view. If all we mean by “The relevant fact has feature X” is “The relevant fact is uniquely represented by the ‘earlier than’ ideology is uniquely represented by the ‘earlier than’ ideology”, then our answer would be question-begging. It would be a mere statement of the view that there are such perspicuity differences, but it will not give us any independent reason to think that such differences are actually there.

Second, even if we do come up with some feature X, we must make sure that this feature will not enter into the metasemantics. For we must still secure the co-reference between the relevant terms. It is exactly at this place that McSweeney’s case of logical constants differ from the case of ‘earlier than’ and ‘later than’. For, assume that facts have their internal conjunctive or disjunctive structure. It seems that those structures will affect the metasemantics of logical notations: it would be natural to assign ‘and’ to facts with conjunctive internal structure, while ‘or’ will be assigned to facts with disjunctive internal structure. To flip that assignment would be pervert. This is not going to be the case with respect to ‘earlier than’ and ‘later than’ given that we want to assign them both to the same range of facts concerning the ordering of events in time.

Finally, we must be cautious of unwanted companies. I take it that there shouldn’t be any difference in fundamentality with respect to typical proper names: let “Eve” and “Fred” be two names for the same electron, it does not really make sense to say that one name as opposed to the other captures the

internal structure of the electron. Now, the perspicuity theorist claims that in cases like ‘earlier than’ and ‘later than’ we are able to make such distinctions. She must, then, explain why the notion of perspicuity finds application in one case but not the other. And I assume that this will be a different task.

The conclusion: while I have not given a conclusive argument against the extra-semantic view on fundamentality, I have at least shown that some possible reasons for supporting it is ineffective or incomplete. The supporter of the extra-semantic view may retreat to primitivism: claiming that there could be fundamentality differences between co-referential terms without providing any theory to explain those differences. But, lapsing into primitivism in the face of challenges is unsatisfying. I submit the best bet is to believe in the cofundamentality thesis and thus the semantically-constrained view.

References

- Cowling, Sam (2013) "Ideological Parsimony", *Synthese* 190 (17), 3889-3908.
- Fine, Kit (2000) "Neutral Relations", *The Philosophical Review*, 109, No. 1, 1-33.
- Lewis, David (1983) "New Work for a Theory of Universals", *Australasian Journal of Philosophy* Vol. 61, No. 4, 343-377.
- Lewis, David (1986) *On the Plurality of Worlds*, Wiley-Blackwell.
- McSweeney, Michaela (2019) "Following Logical Realism Where It Leads", *Philosophical Studies* 176, 117–139.
- Sider, Theodore (2011) *Writing the Book of the World*, Oxford University Press.
- Sider, Theodore (2020) *The Tools of Metaphysics and the Metaphysics of Science*, Oxford University Press.
- Warren, J. (2016) "Sider on the Epistemology of Structure", *Philosophical Studies* 173(9), 2417–2435.

What Wittgensteinian GPTs Can't Do

Filippo Mosca (Rome, Italy)

Jakub Gomułka (Kraków, Poland)

Abstract

In our paper, we focus on using Generative Pre-trained Transformer (GPT) technology, provided by OpenAI as the ChatGPT service, to create customized models that would relate to the philosophical works of Ludwig Wittgenstein. We have created two such models by uploading the content of Wittgenstein's *Tractatus Logico-Philosophicus* and Ms-114 to ChatGPT, accompanied by setting up appropriate instructions. Consequently, we developed interactive chats that facilitate explorations of Wittgenstein's philosophy.

A significant portion of our paper is dedicated to reports from our conversations with these customized GPTs, highlighting both their shortcomings and their potential as research tools. Occasionally, the probabilistic nature of GPT technology leads our models to produce inaccurate or incomplete answers. This inherent unreliability necessitates a cautious approach when using GPTs for research. Scholars should complement GPT insights with traditional methods of investigating the source material. This dual approach maximizes the benefits of GPT technology while mitigating the risks associated with its limitations, ensuring a more robust and reliable research process. This also implies that a language model alone cannot produce a proper philosophical article.

1. The GPT technology

Large Language Models (LLMs) are fundamentally statistical models that predict new tokens (elementary parts of linguistic expressions like words and punctuation marks) based on sequences of previous tokens (Azaria 2022). These models are built on the foundation of artificial neural network technology. At the heart of a neural network are nodes, organized in ordered layers from the input to the output layer, and interconnected to facilitate the propagation of signals. This propagation is achieved through connections called weights or parameters, with each parameter determining the strength of signal transmission from one node to another. During the training phase, weights are adjusted to minimize the discrepancy between the actual output and the desired output. This process is known as machine learning.

The Generative Pre-trained Transformer (GPT) architecture represents a specific type of Large Language Model (LLM) developed by OpenAI since 2018 (Radford et al. 2018). The current versions of GPT are capable of simulating human-like conversations, assisting in text writing (including this one),

correcting their own mistakes, coding functions in various programming languages, and following user-provided examples to complete a range of linguistic tasks without additional training (Zhou et al. 2023). The most advanced version, GPT 4.0, is multimodal, meaning it has the ability to recognize and produce images, graphs, tables, and diagrams of various kinds.

The evolution of GPT technology has experienced remarkable progress across its iterations. It began with GPT-1 in 2018, featuring 117 million parameters and 12 layers, demonstrating the potential of deep learning in text generation. This was followed by GPT-2 in 2019, which had 1.5 billion parameters (10 times more than its predecessor) and 48 layers, bringing improvements in text coherence and contextual awareness. The release of GPT-3 in 2020 marked a significant leap, with 175 billion parameters and 96 layers, greatly enhancing the model's capabilities in generating text across a wider range of subjects and formats without the need for task-specific training. In November 2022, OpenAI made its ChatGPT internet chatbot available to the public, sparking considerable interest with 100 million monthly active visitors within just two months of its launch (Hu 2023). Finally, in March 2023, OpenAI introduced a chat feature based on GPT-4 for ChatGPT users with paid accounts. While the exact number of parameters and other technical details for the latest version of GPT have not been disclosed by the company, it is estimated to possess 1.75 trillion parameters (Gosh 2023).

Information provided by ChatGPT is not necessarily encoded in the network's parameters, as the system can access data from traditional databases when necessary. For instance, if asked for the exact height of Mount Everest, ChatGPT processes this question by searching its knowledge base, where "8,849 meters" is explicitly mentioned. If the answer is not available, such as in the case of certain simple mathematical calculations, ChatGPT 3.5 may provide incorrect information because it attempts to infer the answer from what is scattered throughout the network's parameters (cf. Azaria 2022).

2. Wittgensteinian customized GPTs

A notable feature of the paid version of ChatGPT is the ability to create customized GPT models. This innovation allows for the tailoring of the model to specific needs through establishing specific instructions that is to precede

every prompt given to a customized model. Moreover, the model can be provided with an additional knowledge base uploaded as one or more text files. This additional data can be used by the model to answer discipline-specific questions. A paying ChatGPT user can share their customized model with other paying users or leave it private. It must be mentioned that the customized models run on GPT 3.5 technology, that is, their abilities do not match the GPT 4 advancements.

There are examples of applying this ChatGPT feature in philosophy. For instance, research analyst Kelly Truelove recently developed LucretiusGPT, a model enriched with the text of “On the Nature of Things” by the ancient Roman poet and philosopher Lucretius (Weinberg, 2023). As of February 2024, a significant number of GPT models dedicated to specific philosophers, such as Kant, Heidegger, Hegel, and Schopenhauer, among others, are publicly available. There are also several Wittgensteinian GPTs. It is not possible to immediately assess whether these are made just for entertainment or for scientific purposes. LucretiusGPT, in particular, appears to have been crafted more as an elaborate joke or artistic endeavor rather than as a serious research tool or a means to facilitate learning in Lucretian philosophy (cf. Truelove, 2023).

In our opinion, customized GPTs can be a subject of a genuine scientific investigation. On the one hand, this ChatGPT feature opens a possibility for a significant contribution to the task of interpreting a philosopher by providing a useful tool for research in source-material of a particular thinker. On the other hand, examining interactions between customized models and philosophical texts they are equipped with can yield some insights into the functioning of models themselves.

ChatGPT Tractarian Wittgenstein (TW) was created on the basis of the uploaded knowledge file that included the German text of the *Tractatus Logico-Philosophicus* except for the dedication, the author's Preface, and the footnote that explains the numbering system. It should be underlined that the text is already in public domain and it cannot be a subject to any copyright claims. The file was structured according to JSON dictionary format with numbers of theses as keys and their content as values (more specifically, each value was a tuple that included a full content of a thesis and a list of its sentences;

therefore, all sentences from Tractarian numbered theses appeared twice in the file). The user input was also modified, so to precede each prompt with the following instruction: “The knowledge file uploaded contains the content of Ludwig Wittgenstein's *Tractatus Logico-Philosophicus*. Answer to all the questions as if the file was the primary source of knowledge for you”. TW is a publicly available customized GPT, so anyone who has a paid OpenAI account can query it. The link to the model is available at the WAB site in the section “wittgensteinonline” (<https://wab.uib.no/wittgensteinonline.page>).

Wittgensteinian Oracle: MS 114 (WO) is a customized GPT specifically designed to provide answers based on the content of MS 114. MS 114 is a manuscript that is part of Wittgenstein's *Nachlass*. The knowledge file with which WO is equipped is a JSON file. Each key in the JSON file represents an abbreviation that identifies a specific remark within MS 114, accompanied by the remark's date of composition. These dates may be single, multiple, or even uncertain, as the case may be. Each value is the content of a single remark by Wittgenstein in the manuscript. The model's task is to draw exclusively on this body of knowledge to respond to queries and to maintain a focus on dates of composition (when relevant). The instructions provided are as follows: "Please use the contents of the knowledge file I provide as your primary source for your answers to prompts. This knowledge file (JSON file) has as its keys: the abbreviation name of each individual remark from Wittgenstein's MS 114 in *Nachlass*, accompanied by the date of composition (marked between two parentheses and placed after the word 'date:'). Some examples of abbreviation names: Ms-114,111v[2]et112r[1]; Ms-114,105r[4]et105v[1]; Ms-114,59r[2]; Ms-114,104v[2] Some examples of dates: (date: 1932.05.30), (date: 1933.10.01?-1933.12.31?), (date: 1933.10.01?-1933.12.31?), (date: 1933.10.01?-1933.12.31?). Please note that the date can be multiple or even uncertain (as you can see, there are question marks in that case). Keep this thing about composing dates in mind. I may ask you to respond only by referring to a certain range of time. This JSON file has as values the contents that are to be the primary sources of your answers." Currently there is restricted access to the WO model, because of copyright constraints.

It is important that there are two models rather than just one because they allow us to take into account two very different knowledge base cases. First, the *Tractatus* is a work that Wittgenstein really composed and chose to

publish, while MS 114 is a manuscript in which Wittgenstein jotted down remarks without paying too much attention to a general structure. Secondly, the philosophical style of the thoughts expressed in the two documents is very different because they belong to different stages of Wittgenstein's thought: the logical progression of the “early” Wittgenstein gives way to the descriptive and multifaceted inquiry of the “late” Wittgenstein. Thirdly, there is a notable difference in the prior knowledge of the content of the two datasets: while the standard ChatGPT, especially version 4, possesses some knowledge of the content of the *Tractatus* included in its knowledge base, it definitely lacks prior access to the content of MS 114.

3. Presentation of use cases

Example 1: TW can easily answer quite correctly to any request for explanation of a given Tractarian concept. The following conversation about the concept “Gegenstand” is carried out in German:

Prompt: Was ist ein Gegenstand?

TW: Ein Gegenstand im *Tractatus Logico-Philosophicus* von Wittgenstein ist ein Element der Welt, das in verschiedenen Konfigurationen in Sachverhalten existieren kann. ...

Example 2: TW can easily produce even quite complicated formal expressions, such as those used to define the series of natural numbers in thesis 6.02:

Prompt: Wie definiert der *Tractatus* die natürlichen Zahlen?

TW: Wittgenstein definiert die natürlichen Zahlen im *Tractatus Logico-Philosophicus* durch eine Form der rekursiven Definition mit der Notation von Exponenten einer Operation. Er führt die Zahlen ein durch die Definition $x = \Omega^0 \cdot x$ und $\Omega^v \cdot \Omega^v \cdot x = \Omega^{v+1} \cdot x$, wobei diese Zeichenregeln die Reihe $x, \Omega \cdot x, \Omega \cdot \Omega \cdot x, \Omega \cdot \Omega \cdot \Omega \cdot x, \dots$ als $\Omega^0 \cdot x, \Omega^{0+1} \cdot x, \Omega^{0+1+1} \cdot x, \Omega^{0+1+1+1} \cdot x, \dots$ ausdrücken. Diese Darstellung bildet die Grundlage für seine Definition der Zahlenfolge, wo er beispielsweise die Addition explizit definiert: $0+1=1, 0+1+1=2, 0+1+1+1=3$, und so weiter. Zudem identifiziert er die Zahl als den Exponenten einer Operation, was auf die tiefe logische Struktur hinweist, die Wittgenstein den Zahlen zuschreibt.

Example 3a: When asked (in German) about the general form of truth-function, TW answers correctly, remembering about the notorious bars over variables:

Prompt: Wie kann eine allgemeine Form einer Wahrheitsfunktion definiert werden?

TW: Die allgemeine Form einer Wahrheitsfunktion im *Tractatus Logico-Philosophicus* ist

$[\bar{p}, \bar{\xi}, N(\bar{\xi})]$, und diese wird als die allgemeine Form des Satzes beschrieben ...

Example 3b: However, when asked in English about a summary of the family of thesis 6, it forgets the bars:

Prompt: Can you do the similar thing with the thesis 6 and its family [that is, present a 3-paragraph summary in English]?

TW: Thesis 6 and its related propositions in Ludwig Wittgenstein's *Tractatus Logico-Philosophicus* explore the limits of language, the essence of truth, and the nature of the world in relation to logic and mathematics. The central claim of Thesis 6 is that the general form of a truth-function is given by $[p, \xi, N(\xi)]$...

Example 4a: A series of questions about the relation between the world and reality leading to TW admitting that the world is a part of reality.

Prompt: Ist die Welt ein Teil der Wirklichkeit?

TW: Die Welt ist nicht lediglich ein Teil der Wirklichkeit, sondern sie ist die Gesamtheit der Tatsachen, nicht der Dinge ...

Prompt: Wenn die Wirklichkeit die Gesamtheit der bestehenden und nicht bestehenden Sachverhalte ist und die Welt die Gesamtheit der bestehenden Sachverhalte, warum ist die Welt dann kein Teil der Wirklichkeit?

TW: Die Unklarheit meiner vorherigen Antwort bezüglich der Beziehung zwischen der Welt und der Wirklichkeit im *Tractatus Logico-Philosophicus* bedarf einer Klärung. ...

Daher ist es präziser zu sagen, dass die Welt, verstanden als die Gesamtheit

der bestehenden Sachverhalte, tatsächlich ein Teil der umfassenderen Wirklichkeit ist, die sowohl das ist, was der Fall ist, als auch das, was nicht der Fall ist.

Example 4b: Another conversation on the same topic leading to a different conclusion.

Prompt: Kann man sagen, dass die Welt ist das Bestehen vor Sachverhalten?

TW: Ja, man kann sagen, dass die Welt das Bestehen von Sachverhalten ist.

Prompt: Und, ist die Wirklichkeit das Bestehen und Nichtbestehen von Sachverhalten?

TW: Ja, die Wirklichkeit ist das Bestehen und Nichtbestehen von Sachverhalten.

Prompt: Also, ist die Wirklichkeit die Welt und Nichtbestehen von Sachverhalten?

TW: Die vorherige Information war nicht präzise im Kontext Ihrer letzten Frage ... Die korrekte Darstellung ist, dass die Wirklichkeit das umfassende Konzept ist, das sowohl das Bestehen (die Welt) als auch das Nichtbestehen von Sachverhalten einschließt.

Example 5a: WO has the ability to identify the main topics contained in the knowledge file.

Prompt: Welches sind die wichtigsten Themen?

WO: Die Themen, die in Wittgensteins Manuskript MS 114 behandelt werden, umfassen die Komplexität der Sprache, die Bedeutung und Verwendung von Worten, die Natur von Gedanken und Erwartungen, sowie die Beziehung zwischen Sprache und Realität. Wittgenstein erforscht, wie Sprache die Wirklichkeit abbildet, die Rolle von Definitionen und Regeln in der Sprache, die Unterscheidung zwischen Sinn und Unsinn, und die Konzepte von Befehlen und Handlungen innerhalb sprachlicher Strukturen. Er hinterfragt die Grammatik und die Willkürlichkeit sprachlicher Regeln, sowie die Verbindung zwischen Sprache und psychologischen Mechanismen.

Example 5b: When requested, WO can supply examples of relevant Remarks (along with associated metadata) related to the topic under investigation.

Prompt: 'Die sprachliche Natur von Gedanken'

Kannst du einige Bemerkungen über dieses Thema zitieren?

WO: Hier sind einige Bemerkungen Wittgensteins zur sprachlichen Natur von Gedanken, die in seinem Manuskript MS 114 gefunden wurden:

-Denken als Operieren mit Symbolen: "Wir können sagen: Denken ist das Operieren mit Symbolen. Aber 'Denken' ist ein fließender Begriff, & welcher das 'Operieren mit Symbolen' ist, muss in jedem besondern Fall eigens betrachtet werden. Ich könnte auch sagen: Denken ist Operieren mit der Sprache. Aber 'Sprache' ist wieder ein fließender Begriff" (Ms-114,125r[2]et125v[1], date: 1933.10.01?-1933.12.31?).

-Die Rolle der Sprache im Denkprozess: ...

-Die Unmittelbarkeit des Gedankens: ...

Example 6a: WO can answer negatively to questions regarding the presence or absence of certain topics.

Prompt: Spricht Wittgenstein über die phänomenologische Sprache?

WO: Wittgenstein diskutiert zwar nicht direkt die phänomenologische Sprache in den mir zugänglichen Teilen des Manuskripts MS 114, doch befasst er sich intensiv mit der Bedeutung der Sprache, ihrer Verwendung, den Funktionen von Zeichen und der Beziehung zwischen Sprache und Wirklichkeit. ...

Example 6b: WO can answer positively to questions regarding the presence or absence of certain topics.

Prompt: Spricht Wittgenstein über den Begriff des Verstehens?

WO: Wittgenstein spricht in MS 114 tatsächlich über den Begriff des Verstehens Einige bemerkenswerte Punkte umfassen:

-Die Natur des Verstehens: ...

-Verstehen als psychischer Vorgang: ...

-Verstehen im Kontext von Zeichen: ...

-Die Beziehung zwischen Verstehen und Sprache: ...

-Verstehen als Teil der menschlichen Erfahrung ...

Example 7: WO is sometimes able to answer questions on certain topics, even if the relevant key-words are not explicitly mentioned in the relevant Remarks.

Prompt: Gibt es eine Verständigung, die nicht verbal ist?

WO: Wittgenstein spricht tatsächlich über Formen der Verständigung, die nicht verbal sind. Er erwähnt beispielsweise die Interpretation von Gesten als eine Art der Verständigung, die über die reine Übersetzung in Worte hinausgeht:

"Andererseits sagt man: 'ich verstehe diese Geste', wie 'ich verstehe dieses Thema', 'es sagt mir etwas', & das heißt hier: ich erlebe es, es greift in mich ein. Ich folge ihm mit bestimmtem Erlebnis. Wenn ich, bei irgend einer Gelegenheit, sage: 'ich verstehe diese Geste', meine ich da, daß ich sie in Worten oder andere Zeichen übersetzen kann? Gewiß nicht immer. Ich charakterisiere auch ein Erlebnis. (Die Geste macht einen Eindruck auf mich)."

...

4. Conclusion: Weak and strong points of GPTs

The examples presented in the previous section demonstrate that GPTs excel at answering simple "what" questions. However, their responses are not always entirely accurate, as illustrated by Example 3b, indicating that they cannot be entirely relied upon. For more complex inquiries, such as those concerning the relationship between the world and reality (Examples 4a and 4b), the answers may vary based on the user's approach to the conversation. Additionally, it appears that randomness also influences the responses. Due to space constraints, we did not showcase GPTs' limitations in generating complete lists of items, such as a comprehensive list of all formal concepts or all occurrences of a given concept. It seems that GPTs are better suited to producing examples, as seen in Example 5b, rather than exhaustive enumerations.

Due to the probabilistic nature of this technology, GPTs cannot be utilized in the same way as conventional databases; we cannot expect their answers to be entirely accurate and complete. Nonetheless, they can be valuable when addressing specific research questions. For instance, consider writing a paper on Wittgenstein's concept of "Kalkül." In such a scenario, one could query various GPTs developed for different documents within Wittgenstein's *Nachlass*. Additionally, one could inquire about the context of that concept in particular manuscripts or typescripts. The models might even pinpoint the specific location of discussions related to this concept. However, if there are numerous such locations in a document, one should not anticipate a complete list.

To summarize, GPTs can offer valuable insights for researchers, though they are incapable of producing a complete paper or even a substantial section independently, as the technology is not able to generate texts longer than a few coherent paragraphs that would make sense. Furthermore, due to their inherent limitations, GPTs cannot be entirely relied upon and thus are best utilized in the initial stages of research. This preliminary work must be complemented by further investigation of source material, potentially using other research tools like WAB browsers. Nevertheless, as illustrated by Example 7 in the previous section, GPTs can provide a depth of understanding of document content that significantly exceeds the capabilities of any existing research tool.

References

Azaria, A. (2022) "ChatGPT Usage and Limitations". <https://hal.science/hal-03913837> (Access: February 23, 2024).

Hu, K. (2023) "ChatGPT Sets Record for Fastest-Growing User Base—Analyst Note". <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> (Access: February 22 2024).

Ghosh, L. (2023) "Will Google Gemini outdo GPT-4?" *WIRE19*, <https://wire19.com/google-gemini-vs-gpt-4/> (Access: February 22, 2024).

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018) "Improving language understanding by generative pre-training", *OpenAI Blog*.

Truelove, K (Dec 4, 2023) "On the Nature of Artificial Things", *TrueSciPhi.AI*, <https://www.truesciphi.ai/p/on-the-nature-of-artificial-things> (Access: February 23, 2024).

Weinberg, K. (2023) "Philosophical Uses for LLLMs: Modeling Philosophers", *Daily Nous*. <https://dailynous.com/2023/12/05/philosophical-uses-for-llms-modeling-philosophers> (Access: February 22, 2024).

Zhou J., Ke P., Qiu X., Huang M., Zhang J. (2023) "ChatGPT: potential, prospects, and limitations", *Frontiers of Information Technology & Electronic Engineering*.

Acknowledgement

We are immensely grateful to the University of Bergen and the Wittgenstein Archives for creating excellent conditions for our work. In particular, we want to thank the director of WAB, Alois Pichler, for his invaluable help and support provided to us during the implementation of this project. We warmly thank Tim Grasshöfer for consulting our German queries. Words of thanks are also due to Claus Huitfeldt, Heinz Wilhelm Krüger, Rune Jensen Falch, as well as all other participants of the seminars organized by WAB in February 2024.

Vices and Virtues of Digital Realities: Exploring the Personal And Social Potential of Online Interactions

Carmen Mossner, Sven Walter (Osnabrück, Germany)

Abstract

In the span of merely a few decades, our waking lives have undergone a profound transformation. What was once experienced as corporeal non-digital realities has become ever more supplemented and in part replaced by digital realities. Various studies suggest that this is not always for the better. Our digital realities are said to erode genuine human connections, cause cognitive and emotional deficits and detrimentally affect our mental and physical health. While not unreasonable, this one-sidedly negative depiction of ‘the internet’ misses out on the opportunities and benefits the internet offers to some, despite being detrimental for others.

In what follows we identify structural differences between online and offline realities that are said to have problematic consequences for some people, such as the decrease in emotional and contextual information or the bleak and repetitive nature of the internet. But we also stress that these very same features can harbor profound benefits for others as well, like introverts, elderly or people with autism. We conclude with a meta-analysis that argues that digital realities are not ‘objectively given’ but always construed or enacted by individuals. Consequently, digital realities are not ‘good’ or ‘bad’ *per se*, but only with regard to an individual’s concrete needs and capacities.

Introduction

In the span of merely a few decades, the fabric of human interaction and societal dynamics has undergone a profound transformation. Until very recently, we lived in a corporeal reality that was shaped by face-to-face interactions and thus dependent upon the physical proximity of geographically constrained local communities. In the wake of the digital transformation, this is supplemented, if not replaced, by a mode of existence in *digital realities* that are shaped by and dependent upon global connectivity and instant communication. Not always for the better: As critics keep arguing, digital realities, particularly through and in the context of social media, can erode genuinely human connections, with the shallow and contrived nature of our online identities and interactions causing cognitive and emotional deficits, mental or physical health issues and morally and societally detrimental behavior. While not unreasonable, this one-sidedly negative depiction of ‘the internet’ doesn’t take into account that digital realities also offer *opportunities*. By serving as a lifeline for maintaining relationships across vast distances and facilitating engagement in meaningful global communities, they are not only a

crucial means of self-expression, activism, and exchanging ideas; for many they are also a place that offers resources for cultivating diverse forms of interpersonal relationships.

In what follows, we seek to relativize the skepticism surrounding the digital age. We identify characteristic structural features of online spaces (section 1) that indeed do have problematic consequences for some people (section 2). But we also stress that these very features can go along with profound benefits for others as well (section 3). We conclude with a meta-level analysis that argues that digital realities are not objectively ‘given,’ but always constructed or enacted by individuals, and thus not ‘good’ or ‘bad’ *per se*, but only with regard to an individual’s concrete needs and capacities (section 4).

1. Structural differences between online and offline realities

Interactions in online spaces, be it on social media platforms or via instant messaging services, are prevalent in most of our daily life. We stay in touch with friends that may live next door or across the globe, we relocate job meetings or even psychotherapeutic counselling sessions to video-call platforms, we discuss the next project in the team chat, we play online multiplayer games and we follow the content of friends and strangers on platforms like Instagram and Facebook. At the same time, our corporeal offline interactions dwindle away. Importantly, however, we are not simply witnessing a mere shift of social interactions from offline to online spaces. Given the essentially different environments in which they take place, there is a significant disparity between the structural conditions of offline and online interactions (Mossner and Walter 2024).

Most notably, online communication channels offer less *contextual* and *communicative input*. So far, only some of the modalities present during face-to-face interactions are available online, and oftentimes only a fraction of them are used. For instance, most conversations are based on reading the other’s text messages or listening to their voice messages. Only during video calls are we also able to see their face, some of their facial expressions and their surroundings. And still, only a very limited portion of their surroundings is actually transmitted as noise-cancelling headphones and narrow visual frames filter out large parts that would otherwise contribute to a ‘shared space’ found

in offline interactions (Osler 2021). Not even eye contact can be said to occur in a comparative fashion. We may look the others in the eyes, but since this requires our gaze to be fixated on the screen instead of the camera, it does not appear to the other as through we were actually looking them into the eyes. On top of this, modalities like smell or touch still remain inaccessible to online communication to this day. Similarly, since many of our online interactions are based on text or voice messages instead of real-time video-calls, we often do not receive *immediate feedback* from the person we are interacting with. Sometimes, it can take up to days, weeks or even longer until they respond and the conversation progresses.

Due to this general lack or delay of input, as well as the rather bleak and repetitive nature of the internet as a whole compared to experiences in the non-digital reality, the experiences we make online are experientially and emotionally *less rich*. Emotional expressions, for instance, are conveyed in a more ‘disembodied’ (Fuchs 2014) fashion, since, for example, emoticons or written words largely replace what was otherwise expressed through facial expressions, verbal clues, bodily postures or movements.

Compared to the corporeal world, digital communication channels connect us regardless of geographical locations, enabling interactions that are neither bound to space nor time. This in itself is undoubtedly an amazing achievement. We can easily get and stay in touch with more people including people from different countries, cultures and with different perspectives, which can potentially widen our horizon. And yet, many of our online contacts are kept on a rather superficial level and with a bunch of people we barely know or get to know (e.g., Dutot 2020; Babanazarova 2023). We thus have ever less *background knowledge* about our social contacts, i.e., about their appearance, attitudes, desires, upbringing, and so on.

Adding to this limited background knowledge, we largely move through the internet *anonymously*, making it even harder to establish deeper relationships. Some are less willing to share personal information online, rather staying unknown and unrecognised in the public eye online. Others, however, feel less restraint online precisely because they feel safe behind the curtains of anonymity, making them say things they would not dare say face-to face (Suler 2004).

Lastly, the *amount of information* one is bombarded with online is unprecedented. Especially when frequenting news pages or politically involved platforms, the information can be fairly emotionally triggering. But even outside of those spaces, online users are usually unable to flee or control the amount of emotionally triggering content they are exposed to since advertisements and posts or comments by strangers can be just as emotionally laden and are present on virtually all social platforms. In comparison, the amount of information we are exposed to offline is far more confined, not least because the contacts and interactions we are maintaining offline are comparatively manageable.

There are thus profound differences in what our digital realities afford us during social interactions as compared to offline realities. Over the past two decades, studies have claimed that these differences, among others, lead to tremendous consequences for our social capacities and mental and physical health.

2. The Vices of Digital Realities

The fact that online interactions exhibit such profound structural differences and take up a significant portion of our everyday life nowadays raises the question whether our online realities provide an environment that supports friendly and respectful interactions and support healthy relationships. While even face-to-face interactions can hardly be called empathic and friendly at all times, studies suggest that the differences described above, among others, lead to negative consequences for mental and physical health and social capacities in online users (Mossner and Walter 2024).

One of the apparent downsides of digital communication is the decrease or impediment of social capacities like empathy (e.g., Konrath et al. 2011; Zarins and Konrath 2017). As online interactions are inferior to face-to-face interactions with regard to the amount of personal and contextual information, emotional expressions and modalities they can embrace, they offer us fewer clues about the people we are interacting with, how they are feeling or what goes on in their surroundings. As seen above, we typically do not even receive those clues immediately, as nearly all text-based conversations are steeped in time delays. All of this has been suggested to play

a major role in the decrease in people's interpersonal capacities (e.g., Aagard 2022; Osler 2021) – it just seems to be a lot *harder* to understand or feel empathy towards another person if we receive only limited social clues. A lack of input, background knowledge and feedback can make the few available clues even more arbitrary, in the end, leaving plenty of room for misunderstandings if our assessment of the other person's affective state or behaviour turns out to be faulty. Besides, it has become common knowledge that a frightening number of accounts on platforms like Instagram depict an unrealistically perfect life through methods like stylised or selective posting. Instead of leading to emotional contagion, positive experiences shared online, such as a holiday trip, a date night or newly bought outfit, are suggested to provoke social comparison and jealousy which negatively affect individuals' social skills and mental health, including their empathy, self-esteem and life satisfaction (e.g. de Vries et al. 2018; Gomez et al. 2022; Usán Supervía et al. 2023). And there is much more that can be said about the dynamics on such platforms, including their effects on addictive and compulsive behaviour or feelings of isolation (e.g., Dutot 2020; Babanazarova 2023).

As we move through online spaces largely anonymously, even the most basic similarities between interactants remain hidden. Yet, just like being provided contextual information and perceiving emotional expressions, the *perceived similarity*, encompassing, for example, the recognition of physical similarities or shared perspectives, has shown to be important to facilitate and support social capacities during interactions. For instance, the more similar someone looks to ourselves, the more likely we are to feel empathy towards them (e.g., Heinkes and Louis 2009). Not only does anonymity disconnect us from each other, it also disinhibits our feelings of restraint in terms of what we dare say online, which facilitates behavioural developments like cyber-bullying, hate speech or online radicalisation (e.g., Hangartner et al. 2021; Feddes et al. 2015)

Finally, despite the internet being an incredible source of valuable information, internet users are often bombarded with all kinds of information. Considering the extent of emotionally triggering content, an emotional *numbness* can set in eventually. We may thus develop an insensitivity to new information which might negatively affect our personal and emotional relationships. Perhaps, we are not only less sensitive to relevant emotional clues but also reduce the amount of emotional expressions we show.

Although far more might be said about the detrimental effects of the structural differences between online and offline spaces, the aspects described so far are the most relevant for our purposes: While they pose challenges to many people during social interactions online, they might actually be precisely the reason why social interactions work better or more easily for others. In the following section, we claim that humans experience online environments in partly fundamentally different ways, allowing some to flourish during social interactions in their own digital realities while others are faced with the challenges described above.

3. The Virtues of Digital Realities

The picture painted in section 2, plausible as it may sound, is too one-sided, a caricature of ‘the’ internet that doesn’t do justice to the rich variability of the platforms, spaces, and communities through which we engage with one another online. Posting a stylised selfie on Instagram is not the same thing as speaking candidly in a closed forum for recovering addicts, say, or find a community of other supportive autistic gamers in a massively multiplayer online role-playing game, where one can be and do things one can’t do in the ‘real’ offline world predominantly shaped by ‘neurotypical’ people. In fact, it is precisely the structural differences between online and offline spaces regarding, among other things, suitable background knowledge, a perceived similarity, the quality and quantity of the input, the immediacy of the feedback etc. that happen to negatively affect people’s psychological well-being that also open up the possibility for others to benefit from and flourish in the online realities in which they live.

In our digital realities we typically certainly do have less background knowledge about many of the people we interact with, with the result that the shared perspective that fosters mutual understanding and empathy is less likely. Yet, people with niche-interests, be they determined by their physiological conditions (e.g., people with rare diseases), a matter of cultural orientation (e.g., people with atypical sexual preferences) or mere hobby horses (e.g., people collecting rare memorabilia) can find likeminded contacts with which they share interests, experiences and attitudes more easily and readily online than in the offline world.

In our digital realities, we are typically certainly faced with a kind of disinhibiting anonymity that makes people more likely to say things they would not dare say in 'real' life, fostering hate speech, cyber-bullying and, eventually, aggression. Yet, members of the LGBTQ+ community, alcoholics or other social minorities can meet in forums, exchange their often shared experiences and establish feelings of belonging, being understood etc. – something that many members of certain social groups, including sick and disabled people, do not receive from their friends and family outside of their digital realities.

In our digital realities, communications typically certainly lack the guarantee of immediate feedback and other essential characteristics that make conversations smooth, like turn-taking signals and other information provided by the broader conversational context, that makes people more likely to misunderstand each other or feel hurt or at least more difficult to decipher the others' intentions. And yet, again, it is precisely this that can make people feel more comfortable and safe in their online realities. First, autists or socially anxious people, for instance, benefit from no immediate contact and from having the time to interpret and prepare their response and might use the extra time offered by the possibility of a delayed response to process information in the way they need to make an empathic response. Second, people with bodily restrictions such as people with hearing aids or tinnitus, may experience non-verbal online communications as more pleasing than offline ones, and the same might be true for, say, stutterers. Third, some people have untypical behavioral characteristics (very touchy people, no feeling for turn-taking etc.), due to which they might experience rejection during offline interactions but with which they need not deal with online. Fourth, it is precisely the potentially ambiguous nature of online conversations that have lead to new 'expressions' that require much less context for their correct interpretation. A crying emoticon, for instance, can make deciphering the other's affective state easier for everyone, potentially, but especially for people like autists that characteristically have trouble interpreting contextual clues.

In our digital realities, communications typically certainly also hold the danger of a general 'emotional numbness,' given that the emotions communicated there are often 'disembodied', i.e., expressed through written words and/or emoticons only, without facial expressions, bodily postures etc. But, first,

introverts or socially anxious people can benefit from the disembodied communication channels, because they do not have to deal with the quantitatively rich social face-to-face input (e.g., Valkenburg and Peter 2009). Second, while quantitatively less rich is bad for those for whom qualitatively rich is or used to be the norm, it might be an improvement for those who didn't have or otherwise wouldn't have any social interactions and communications at all: For elderly people who cannot go out, the online input in a digital reality shared with their remote and scattered families might be the only way to get any social input at all.

In our digital realities, we are typically bombarded with information 24/7, constantly emotionally triggered by largely anonymous people, leading to an emotional exhaustion. For some, however, precisely the constant availability of others with a shared background, anonymous as they otherwise might be, might be what makes all the difference: An alcoholic who had a really bad day and is on the verge of relapse, sitting in front of the bottle, doesn't have the time to wait for the next AA meeting three weeks from now, but needs to talk to people who know what they are going through in the moment.

These are cases where the bad isn't necessarily bad since what might be bad for some, presents itself beneficial for others. This is one way in which a purely negatively focused take on 'the' internet can be put into perspective. Another way is actually to grant that a feature affects indeed all people, but to deny that it is necessarily bad. Consider for instance, the general emotional numbness triggered by online realities and the prevailing of social comparison over emotional contagion: We no longer feel sad when other people are feeling sad. However, can we hope to cope with the challenges of our radically altered environment if we continue to chase the sort of empathy that results from mechanisms that have evolved under completely different circumstances? Is greater 'genuine' empathy really what we need (Mezzenzana and Peluso 2023)?

4. Conclusion

Our exploration of the effects of online spaces highlights the nuanced impact of aspects of the internet on individuals, which have widely been regarded negative. While structural differences between online and offline interactions, such as the lack of contextual and conversational input, the absence of

immediate feedback or the superficial nature of contacts and conversations, can indeed have detrimental effects on some people, as proclaimed in various studies, they can prove highly beneficial for others, for instance for people with special interests or habits, people with autism, introverts, elderly or socially anxious people. Thus, it is crucial to recognize the internet not as a homogeneous space or an objectively 'given' reality, but rather as a complex environment comprising a myriad of differently experienced realities. Just as people vary in their abilities, experiences, behaviors and needs, so do the effects of various aspects of the internet on them. Taking on a more differentiated view on the internet by acknowledging this diversity is essential when designing or modifying online spaces in ways beneficial for the people inhabiting those spaces.

References

Aagaard J (2022) "On the dynamics of Zoom fatigue", *Converg* (28)6:1878–1891. <https://doi.org/10.1177/13548565221099711>

Babanazarova, M. (2023) "The Detrimental Effects of the Internet in Today's World: A Comprehensive Analysis", *American Journal of Language, Literacy and Learning in STEM Education* (2993-2769), 1(6), 174-176.

de Vries DA, Möller AM., Wieringa MS., Eigenraam AW., Hamelink K. (2018) "Social comparison as the thief of joy: Emotional consequences of viewing strangers' Instagram posts", *Media Psychol* 21(2), 222–245. <https://doi.org/10.1080/15213269.2016.1267647>

Feddes AR, Mann L, Doosje B (2015) "Increasing self-esteem and empathy to prevent violent radicalization: a longitudinal quantitative evaluation of a resilience training focused on adolescents with a dual identity", *J Appl Soc Psychol* 45(7), 400–411. <https://doi.org/10.1111/jasp.12307>

Fuchs T (2014) "The virtual other: Empathy in the age of virtuality", *J Conscious Stud* 21(5–6), 152–173.

Gomez M, Klare D, Ceballos N, Dailey S, Kaiser S, Howard K (2022) "Do you dare to compare?: The key characteristics of social media users who frequently make online upward social comparisons", *Int J Hum Comput Interact* 38(10), 938–948. <https://doi.org/10.1080/10447318.2021.1976510>

Hangartner D, Gennaro G, Alasiri S, Bahrnich N, Bornhoft A, Boucher J, Demirci BB, Derksen L, Hall A, Jochum M, Murias Munoz M, Richter M, Vogel F, Wittwer S, Wüthrich F, Gilardi F,

Donnay K (2021) "Empathy-based counterspeech can reduce racist hate speech in a social media field experiment", *Proc Natl Acad Sci* 118(50), e2116310118. <https://doi.org/10.1073/pnas.2116310118>

Heinke MS, Louis WR (2009) "Cultural background and individualistic-collectivistic values in relation to similarity, perspective taking, and empathy", *J Appl Soc Psychol* 39(11), 2570–2590. <https://doi.org/10.1111/j.1559-1816.2009.00538.x>

Konrath SH, O'Brien EH, Hsing C (2011) "Changes in dispositional empathy in American college students over time: a meta-analysis", *Pers Soc Psychol Rev* 15(2):180–198. <https://doi.org/10.1177/1088868310377395>.

Mezzenzana F, Peluso D (eds) (2023) *Conversations on empathy: Interdisciplinary perspectives on imagination and radical othering*. Routledge, London.

Osler L (2021) "Taking empathy online", *Inq*, 1–28. <https://doi.org/10.1080/0020174x.2021.1899045>

Suler J (2004) "The online disinhibition effect", *Cyberpsychol Behav* 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>

Usán Supervía P, Salavera Bordás C, Quílez Robres A, Lozano Blasco R, Latorre Coscolluela C (2023) "Empathy, self-esteem and satisfaction with life in adolescent", *Child Youth Serv Rev*, 144.

Valkenburg, P. M., & Peter, J. (2009) "Social consequences of the Internet for adolescents: A decade of research", *Current directions in psychological science*, 18(1), 1-5.

Zarins, S, Konrath, S (2017) "Changes over time in compassion-related variables in the United States" In Seppälä EM, Simon-Thomas E, Brown SL, Worline MC, Cameron CD, Doty JR (Eds.), *The Oxford handbook of compassion science*. Oxford/New York: Oxford University Press, 331–352.

Why Structuralism Cannot Defeat Scepticism

Alex Murphy (London, UK)

Abstract

David Chalmers (2018) gives an anti-sceptical argument that invites us to reconsider the relationship we have to reality. He argues that I cannot rule out nightmarish Cartesian scenarios, such as the brain in vat scenario or evil demon hypothesis. Yet, he continues, despite this, such scenarios do not threaten the truth of my beliefs, or indeed my knowledge, about physical reality. This is because structuralism is true – i.e. physical claims are structural claims. They are made true whenever the relevant structure obtains. Furthermore, he argues, that structure obtains in most Cartesian scenarios. So, on such scenarios my physical claims are broadly true. Things may not be how they appear, but I am mostly right about them – whatever the sceptical weather.

In response, I present three objections. First, I argue that Chalmers cannot coherently present the relevant kinds of Cartesian scenarios – they are therefore not first-person epistemic possibilities as he suggests. Second, I argue that his attempts to overcome Newman's Problem fail and that even if they succeeded, they would threaten both the major motivation for his view and the scope of the anti-sceptical result. Third, I present a counter-example Cartesian scenario on which there are no unique realizers for our physical property roles as Chalmers requires there be.

Chalmers attempts to use considerations about how linguistic content is fixed to overcome scepticism. I show that such an attempt fails, and along the way I highlight important lessons about our language and our relationship to external reality.

1. Chalmers' anti-sceptical argument

Chalmers' response to the sceptic focuses on the truth of our beliefs (rather than their justification or with knowledge). He endorses *veridicalism* – on sceptical scenarios our claims are mostly true:

Even if the evil-demon hypothesis is true, I still have hands. (Chalmers 2018: 626)

Call any claim that a particular Cartesian scenario obtains a *Cartesian hypothesis*. Call global sceptical scenarios where the subject's experiences is qualitatively the same, whilst the environment around them globally changes a *Cartesian scenario* – e.g. the brain in a vat (BIV) scenario in which I am (and always have been) a disembodied BIV.

Chalmers understands these as first-person epistemically possible (cannot be ruled out a priori) ways our world could be (Chalmers 2018: 632 & 2012 E10).

So the question is: if BIV scenario holds, do I have hands? (Chalmers 2018: 633) These hypotheses are about my situation, not someone else's, nor about what *would be* so, *were* I to be in the situation.

Chalmers' anti-sceptical argument:

- (1) In explanatory Cartesian scenarios, many of our physical beliefs are true.
- (2) Non-explanatory Cartesian scenarios are ruled out by abduction.
- (3) [So,] All Cartesian scenarios are either ruled out by abduction or are scenarios where most of our physical beliefs are true. (Chalmers 2018: 649)

(1) is supported by Chalmers' structuralism – the view that physical claims are equivalent to structural claims. Structural claims are claims using only logical and mathematical expressions (alongside limited auxiliary expressions) (Chalmers 2018: 634-5). Chalmers holds that the meaning of physical terms is given by a role extracted from our folk physical theory (Lewis 1966). The referent of those terms is the realizer of that role.

Chalmers' motivates this view via Twin Earth cases such as:

El Greco World: Just like our world, except things are elongated 2:1 in a fixed direction...On El Greco World, Max's counterpart Twin-Max is an elongated functional duplicate of Max. On Earth, Max sees a square and says 'That's square.' Twin-Max sees (what we'd call) a 2:1 rectangle and says 'That's square.' (Chalmers 2019b: 124)

Chalmers thinks both Max and Twin-Max are correct. Their terms 'square' refer to different things (Chalmers 2019b: 125). In other words, 'square' is Twin-Earthable. He takes this to evidence the structuralist claim that the referent of 'square' is whatever the plays the *square-role* in our physical theory. For Max that's one thing, for Twin-Max it's another. What goes for 'square', Chalmers thinks, goes for all physical terms.

If structuralism is correct, the reference of our physical terms varies systematically between Cartesian scenarios. We refer to whatever fulfils the relevant structural role in the scenario. These realizers differ in each scenario,

so our terms refer differently in each scenario. This systematic variance in reference secures the truth of our claims across scenarios.

2. Chalmers' incoherent set-up

This, of course, assumes that we Cartesian scenarios are coherent first-person epistemic possibilities. Chalmers' central claim is:

Even if the evil-demon hypothesis is true, I still have hands. (Chalmers 2018: 626)

Now consider substituting in the BIV hypothesis:

Even if the BIV hypothesis is true, I still have hands.

Since BIVs are disembodied, Chalmers is seemingly committed to:

Even if I have no hands, I still have hands.

This is contradictory. Note, nothing is special about 'hands' or hands, the problem is systematic. Chalmers' anti-sceptical result is, supposedly, achieved because the reference of physical terms is fixed structurally. This allows terms like 'hands' to refer differently across scenarios. Yet for Chalmers to present Cartesian scenarios, physical terms must refer as they do in the non-Cartesian scenario.

Chalmers asks me to imagine a first-person epistemic possibility wherein I am a handless BIV. He tells me that since the reference of my physical vocabulary is fixed structurally, in this scenario my term 'hands' refers to computer code (Putnam 1981: 14). If so, then "I have hands" is a true claim about computer code. Yet, then "I am a BIV without hands" is a false claim about computer code. Crucially, I cannot both state the first-person epistemic possibility that I am a handless BIV and claim I have hands. Yet this is what Chalmers suggests.

To deliver veridicalism, one must give a semantics on which the reference of terms like 'hands' systematically vary depending on what Cartesian scenario obtains. Chalmers does so. Unfortunately, this leaves no room for the

statement of the Cartesian hypotheses. The reference of all physical terms will have shifted – including those used to formulate the Cartesian hypothesis! Veridicalism leads to the conclusion that these scenarios aren't possibilities.

3. Newman Bites Again

My next objection concerns Chalmers' structuralism: the view that our physical claims are structural claims. Chalmers' structuralism forms a premise in a sub-argument designed to support:

- (1) In explanatory Cartesian scenarios, many of our physical beliefs are true.

The sub-argument:

- (A) All physical claims P are equivalent to structural claims S(P).
- (B) For all positive physical claims P, if S(P) is true in a non-Cartesian scenario, it is true in any corresponding Cartesian scenario.
- (C) [So,] For all positive physical claims P, if P is true in a non-Cartesian scenario, it is true in any corresponding Cartesian scenario. (Chalmers 2018: 640)

Premise (A) is Chalmers' structuralism, it uses the Ramsey-Carnap-Lewis method. The details won't concern us but note it exclusively and exhaustively divides troublesome (t-terms) from okay terms (o-terms). O-terms are assumed to have an uncontentious interpretation (Button&Walsh 2018: 55). The method allows us to give an interpretation of our t-terms by providing a uniquely-fulfilled reference condition using o-terms.

Chalmers suggests we apply the method to physical theory (T) - e.g. the t-term 'charge' refers to whatever plays the role of charge in T. Thus, we get a definition of 'charge' using only logical and o-terms. The property of charge is the unique realizer of the charge role. Chalmers uses these definitions of t-terms to translate physical sentences into structural sentences.

Unfortunately, logical structure alone doesn't uniquely pick out the intended t-term referents. This is Newman's Problem. Newman writes, for any structure W:

Any collection of things can be organised...to have the structure W, provided there are the right number of them. (Newman 1928, 144)

Structural information determines the cardinality of the domain but not the referent of t-terms. Chalmers, like most structuralists, responds by bolstering the o-terms until he can rule out deviant interpretations of the t-terms.

Chalmers first bolsters the o-terms with casual (and nomic) terms (Chalmers 2018: 638). By assuming causal/nomic terms are o-terms, Chalmers hopes to rule out deviant t-term interpretations on causal grounds. Unfortunately, things aren't so simple. I'll only discuss causal terms, but the same problems arise with nomic terms.

To illustrate, briefly assume that the o-term/t-term divide separates the observational and theoretical. Now, consider the relation 'x causes y.' There will be, across the divide, various instances of this relation. Claims about microscopic chemical reactions serve as theoretical examples. Macroscopic causal claims serve as observational examples. What about 'atmospheric pressure caused the barometer mercury to move'?

The structuralist, insisting on the o/t-term divide, must split the mixed instances into the o-term/t-term camps, using a different predicate for each (Button&Walsh 2018: 67). Once split, we cannot know the intended interpretation of the predicate on the t-term side. This causal vocabulary will be just more t-vocabulary whose interpretation is at question.

We cannot just assume that causal terms on the t-side have the same interpretation as those on the o-side. As Buton & Walsh write, where P is an unintended interpretation:

it is profoundly unclear...that we have sufficient handle on the 'troublesome side' of causation, even to articulate the idea that P has failed to respect the (troublesome) causal relationship between troublesome vocabulary and troublesome entities. (Button & Walsh 2018, 67)

This is a general reason to think the response fails. Furthermore, Chalmers in particular should avoid it for two reasons. First, by his own lights causal terms should be t-terms because they're Twin-Earthable. If Twin-Earth experiments motivate treating physical terms as t-terms, then they do so for causal terms. Consider:

Common-Cause Twin-Earth: Exactly like Earth except every apparent case of causation is an instance of common causation. Though a brick seemingly causes a window to smash, in fact, it doesn't. Rather, some prior cause is responsible for the occurrence of both the brick's movement and the window's smashing.

Imagine Twin-Carol on Twin-Earth, a replica of Earth-Carol. I contend that Carol and Twin-Carol refer to different things with their utterances of 'cause'. This would be evidence that 'cause' is Twin-Earth-able (analogous to Chalmers' *El Greco*). So, Chalmers ought to treat 'cause', like 'square', as a t-term.

One might worry that Common-Cause Twin-Earth isn't possible but consider the following Cartesian scenario as an illustration:

Common-Cause Simulation Scenario: The world we experience is a simulation run on a base-level reality computer. Apparent physical causation is common causation. Digital phenomena are causally responsible for all events despite how experience presents it. A brick seemingly causes a window to smash but actually a base-level computer data structure is a common cause.

In this case, the digital processes carry the causal weight. The brick doesn't cause anything. A speaker in the simulation uttering 'cause' isn't referring to causation as it appears at base-level reality. Rather, she is referring to common causation. As Putnam notes:

Brains in a vat can no more refer to what the unenvatted call 'causation' than they can to what the unenvatted call 'fire'. (Putnam 1992: 362)

I think causal terms are Twin-Earthable. That said, *only this weaker claim is necessary*: if El Greco world motivates structuralism, then this motivates a similar view for causal terms. Causal terms are *as Twin-Earthable as* physical terms. To the extent Chalmers has a reason to treat physical terms as t-terms,

he has a reason to treat causal terms similarly. The motivation for structuralism is now a reason against using causal terms to resolve Newman's Problem.

There's another reason Chalmers shouldn't turn to causal terms. His structuralism supports the anti-sceptical argument by allowing reference to systematically shift across Cartesian scenarios. If Chalmers makes causal terms o-terms, then their reference won't be fixed structurally and so won't vary across Cartesian scenarios. Depending on the scenario, our causal claims would then be false. Designating causal terms as o-terms might help with Newman's Problem but it invites scepticism about causal claims. This is disastrous considering how prevalent causal terms are. Not only is explicit causal language common in physical talk, but there are also causal entailments for many verbs – e.g. 'to eat'. Causal language is implicit in lots of talk.

Chalmers tried to overcome Newman's Problem by treating causal/nomic terms as o-terms. I've only discussed causal terms but I contend that the same objections hold for nomic terms.

Chalmers' second attempt to overcome Newman's Problem is add the second-order predicate 'natural' (expresses naturalness of first-order properties) to the o-terms (Chalmers 2018: 638). Chalmers claims that naturalness is a structural notion – i.e. not logical, but appropriate in purely structural descriptions (Chalmers 2022, 101).

It's unclear why naturalness is more of an objective notion than squareness, distance, or charge. Chalmers needs an argument – i.e. we need a reason to think that 'natural' is an o-term. Unfortunately, the same problems arise as with causal terms. I won't rehearse each point, but a few notes are important.

First, one can Twin-Earth 'natural':

Unnatural Twin-Earth: Twin-Earth is different to Earth in just one respect. There are no natural properties on Twin-Earth, only disjunctive properties. The causes of all experiences are wildly disjunctive.

To further illustrate we could imagine a simulation scenario but to avoid introducing too much from Chalmers' virtual metaphysics, consider (Chalmers 2017, 2019a):

Unnatural Demon Scenario: The world I experience is concocted by an evil demon who either biochemically manipulates my brain or programmes a supercomputer to electronically stimulate my brain. The causal basis of my experiences is a matter of biochemistry and digital programming in equal measure.

Imagine metaphysicians in the scenario who predicate properties with 'natural' because of how they feature in phenomenal experience and physical explanation. Intuitively, they don't refer to naturalness with 'natural'. The properties they so predicate will be disjunctive (between biochemical and digital realizers) and so unnatural. Their term 'natural' doesn't refer like ours does - 'natural' is Twin-Earthable.

The same problems arise. If El Greco motivates structuralism, then Unnatural Demon Scenario motivates treating 'natural' as a t-term. Additionally, treating 'natural' as an o-term leaves our naturalness claims open to doubt. Note, that since our ordinary claims aren't often about naturalness this consequence is less problematic for Chalmers' overall anti-sceptical result.

Overall, Chalmers' appeals to causal/nomic relations, and naturalness, don't solve Newman's Problem.

4. Objection 3: Counterexample: Unique Realizers and Abduction

The Ramsey-Carnap-Lewis method requires the unique satisfaction of t-terms. Securing the anti-sceptical result requires claiming that our t-terms are uniquely satisfied on all Cartesian scenarios. Unnatural Demon Scenario is a counterexample. On this scenario there are two equally good candidate realizers for each physical property role.

Note, he cannot claim that each role is uniquely realized by the disjunctive property which combines the relevant biochemical and digital realizers. This sort of deviant property is meant to be ruled out by his appeal to naturalness. Chalmers needs to rule out the scenario. I'll consider two options.

First, given his other comments, he might claim the scenario doesn't meet a necessary condition for being a Cartesian scenario – sharing causal structure with the ordinary scenario. By 'causal structure', I take it, Chalmers means the

reified collection of the causal relations among entities. Two scenarios to share causal structure iff they feature the same number of entities and those entities are structurally isomorphic with regards their causal relations. Chalmers writes:

for the evil-demon scenario: supposing the demon is modelling classical physics, then for every classical particle there will presumably be some sort of representation in the demon's mind (Chalmers 2018: 642)

Furthermore:

If two properties have a systematic causal connection tying them together in the actual world, the corresponding virtual properties will have a systematic causal structure in the simulation. (Chalmers 2018: 642)

In fact, Chalmers only requires that Cartesian scenarios have *at least* the causal structure of their non-Cartesian counterpart. His argument only aims to secure the truth of our positive physical claims, which are “intuitively one[s] that cannot conceivably be rendered false by adding something to a world” – e.g. “There are more than five particles.” (Chalmers 2012, 111) Examples of negative sentences include “There is no nonphysical ectoplasm” and “All life is made of DNA.” (Chalmers 2012, 112) Positive physical claims, then, are claims about this minimal shared causal structure. Such claims are true on Cartesian scenarios (according to Chalmers). Cartesian scenarios may have extra structure but claims about this structure are not immune from scepticism.

So, Chalmers may argue that Unnatural Demon Scenario has more causal structure than the ordinary scenario, pointing to the alternating causal bases. Any claims about that extra causal structure aren't secured by his anti-sceptical argument. So, despite appearances, Unnatural Demon Scenario isn't a counterexample. The relevant claims are false on that scenario, and always were for Chalmers.

First, this response underestimates the generality of the problem. Every property in the scenario will be realized disjunctively. All physical claims are at risk. Second, the positive/negative claim distinction doesn't track anything epistemologically or doxologically interesting enough for it to determine what is(n't) vulnerable to scepticism. If any claim can be secured by an anti-global-

sceptical argument, the following can: “I don’t have 50ft tentacles.” Yet, it’s a negative claim. Its truth in Cartesian scenarios is not secured by Chalmers’ argument. Chalmers’ anti-sceptical argument has a surprisingly restricted scope.

Chalmers’ might alternatively try to rule out Unnatural Demon Scenario using abduction – see his (2):

(2) Non-explanatory Cartesian scenarios are ruled out by abduction.
(Chalmers 2018, 649)

For Chalmers, non-explanatory scenarios are those which do not give a systematic explanation for the regularity of our experience – e.g. a scenario whereby our experiences are caused by random fluctuations of matter. If I’m in a non-explanatory scenario, the relevant causal structure won’t be present and so my physical beliefs will be false. Securing the truth of our physical claims requires consistent causes for our experiences (Chalmers 2018: 648). Chalmers thinks we can reject non-explanatory scenarios on abductive grounds “precisely because these scenarios have no good explanation of our experiences, and many better explanations are available.” (Chalmers 2018: 648)

Unnatural Demon Scenario is not a non-explanatory scenario, however. It gives a reasonable explanation of our experiences. Note, ‘reasonable’ is best read as ‘reasonable given the dialectic.’ Unnatural Demon Scenario is relevantly like classic Cartesian scenarios. So, although it would be unreasonable to raise when discussing applied ethics, it’s reasonable to raise when discussing scepticism.

Chalmers may object that Unnatural Simulation Scenario doesn’t explain the regularity of our experiences, at least not in the simplest manner – there is no explanation for why biochemical and digital mechanisms cause similar experiences.

We should consider what standard of abductive reasoning is appropriate. Chalmers himself recognises that abductive responses to scepticism require an inappropriately strong use of abduction (Chalmers 2018: 649). If Chalmers abductively rules out Unnatural Demon Scenario, then he will be making an

inappropriately strong abductive move too. It would amount to the implausible claim that it's more likely that there is a unique natural referent for every t-term than not:

[if] there is some reasonable explanation, there must be relevant structure in the environment. Then much of the structure required for the truth of our physical beliefs will be present, and many of those beliefs will be true. (Chalmers 2018: 648)

If our t-terms (physical terms) aren't uniquely realized, then don't refer, and our physical beliefs are false/meaningless. So, in effect, he's claiming that any scenario that provides a reasonable explanation of our experiences has a unique realizer for all physical property roles. This is unsupported.

Claiming that Unnatural Demon Scenario isn't a reasonable explanation of our experiences because it doesn't give a simple enough explanation of their regularity is claiming that a scenario with unique realizers for each role is more likely than one without. This is the kind of work that Chalmers himself recognises abduction cannot do.

We cannot rule out Cartesian scenarios by claiming they're less likely than the reality of the world as it appears, as Chalmers admits. Similarly, we cannot rule out one Cartesian scenario rather than another based on the complexity of the causal origins of our experiences. However likely Unnatural Demon Scenario is, it's no less likely than Descartes' original scenario. Our priors for it obtaining shouldn't be lower than those regarding Descartes' original scenario. Abduction cannot rule it out.

5. Conclusion

Chalmers' anti-sceptical argument has three problems. First, on Chalmers' set-up there's no coherent way to present these Cartesian scenarios. Second, Chalmers' solutions to Newman's Problem fail and even if they succeeded, they would undermine the major motivation for his view and restrict the scope of his anti-sceptical result. Third, there's a counterexample to Chalmers' central claim that Cartesian scenarios uniquely realize our physical terms. Structuralism cannot defeat scepticism.

References

- Button, T. (2013) *Limits of Realism* (Oxford: Oxford University Press).
- Button, T. & Walsh, S. (2018) *Philosophy and Model Theory* (Oxford: Oxford University Press).
- Chalmers, D. (2012) *Constructing the World* (Oxford: Oxford University Press).
- Chalmers, D. (2017) "The Virtual and the Real", *Disputatio*, 9(46), 309-352.
- Chalmers, D. (2018) "Structuralism as a Response to Skepticism", *Journal of Philosophy*, 115(12), 625-660.
- Chalmers, D. (2019a) "The Virtual as the Digital", *Disputatio*, 11(55), 453–486.
- Chalmers, D. (2019b) "Three puzzles about spatial experience", in: Pautz, A. & Stoljar, D. (eds.) *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness* (Cambridge MA: MIT Press), 109-137.
- Chalmers, D. (2022) *Reality+: Virtual Worlds and the Problems of Philosophy* (US – New York: W. W. Norton) and (UK – London: Penguin).
- Lewis, D. (1966) "An Argument for the Identity Theory", *Journal of Philosophy*, 63, 17–25.
- Putnam, H. (1981) *Reason, Truth, and History* (Cambridge: Cambridge University Press).
- Putnam, H. (1992) "Replies", *Philosophical Topics*, 20(1), 347–408.
- Newman, M. (1928) "Mr. Russell's Causal Theory of Perception", *Mind* 37(146), 26-43.

A Wittgensteinian Perspective on Realism/Anti-Realism Debates

Ragna Talea Oeynhausen (Munich, Germany)

Abstract

Analytic ontology is dominated by the debate between realists and anti-realists about what exists. The focus lies on certain kinds of entities, such as mathematical entities or properties, and their existence. This paper, however, will not add to that discussion. Instead, I aim to highlight some of the underlying presuppositions of realism/anti-realism debates.

Drawing on Wittgenstein's theory of language-games, this paper will challenge the idea that existence is a univocal concept. I will argue that "exist" is used in a number of ways and thus varies in meaning. In particular, I will criticize the identification of existence with existential quantification. I will then suggest that realists and anti-realists engage in different language-games and use different concepts of existence. In order to avoid verbal disputes, it follows that any debate about what exists must be preceded by a discussion of the presupposed concept of existence.

Ultimately, I propose that (non-verbal) disagreement about existence may highlight a difference in understanding of the nature of the entity in question. This brings about a shift from considerations of existence to considerations of essence. By doing so, I hope to emphasize not only the complexity of ontological debates, but also the interconnectedness of existence and essence.

1. Introduction

In this paper, I develop a Wittgensteinian perspective on the contemporary debate in analytic ontology. Specifically, I am concerned with debates in which two parties disagree about what exists (e.g., realism vs. anti-realism about mathematical entities). Building on Wittgenstein's theory of language-games proposed in the *Philosophical Investigations*, I argue that existence claims are made in multiple (possibly distinct) language-games. It follows that the meaning of "exist" varies between language-games. Ultimately, I hope not only to eliminate linguistic confusion, but also to highlight the multiplicity of assumptions that underlie any talk of existence.

2. The Realism/Anti-Realism Debate

Mathematical realists disagree with mathematical anti-realists about whether mathematical entities such as numbers or sets exist. But ontological debates are not confined to mathematics. More generally, there is disagreement about the existence of abstract objects, properties, or fictional characters. Roughly

speaking, the anti-realist will reject the existence of the entities in question, while the realist will grant it. Despite much discussion, the debate has not been settled in favor of either party. Instead, more and more concerns have arisen about whether the realist and the anti-realist disagree on a substantial level; or whether such debates boil down to verbal disputes about how language should be used. The idea that ontological disputes are merely a matter of linguistic preference goes back to Carnap (1956). More recently, his ideas have been taken up by Hirsch (2002, 2009) and Thomasson (2015, 2017) to advance a deflationary view of analytic ontology.

I will not attempt to settle this debate here. Instead of arguing for one side or the other, I will take a step back and consider the meaning of the word "exist" in existential claims. Clearly, the meaning of existence must be settled *before* it can be decided whether a particular entity exists or not. Since our intuitive agreement with existential claims such as "there are prime numbers greater than one hundred" and "there are trees in the garden" is central to realism/anti-realism debates, these claims will be the focus of the present discussion. In particular, I hope to account for their intuitive truth, while acknowledging that there may be reasons to deny the existence of trees or prime numbers in other contexts.

3. A Brief Sketch of Wittgenstein's Philosophy

Before I turn to an analysis of the realism/anti-realism debate, it will be necessary to briefly outline Wittgenstein's conception of language. The late Wittgenstein's work is characterized by a break with the theory of language of the *Tractatus*. Wittgenstein rejects the reference model of meaning, i.e. the idea that the meaning of a word is given by some object that the word picks out, and moves to a conception of language as a *practice*. The use of an expression within the linguistic practice then emerges as constitutive of its meaning. "For a *large* class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: the meaning of a word is its use in the language." (PI 2009 § 43; emphasis in original) In particular, the meaning of an expression cannot be determined in isolation. One must examine the practice in which the speaker is engaged in order to fully grasp what she means to say.

Moreover, it is Wittgenstein's central observation that there is no single use of language. On the contrary, language is used in a number of different ways.

Think of the tools in a toolbox: there is a hammer, pliers, a saw, a screwdriver, a rule, a glue-pot, glue, nails and screws.—*The functions of words are as diverse as the functions of these objects.* (And in both cases there are similarities.) (PI 2009 § 11; emphasis mine)

But how many kinds of sentence are there? Say assertion, question and command?—There are *countless* kinds; countless different kinds of use of all the things we call 'signs', 'words', 'sentences'. And this diversity is not something fixed, given once for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten. (PI 2009 § 23; emphasis in the original)

Wittgenstein introduces the notion of a "language-game" precisely to capture the multifaceted, dynamic nature of language. Instead of speaking of language as a whole, he speaks of language-games, each of which reflects a particular use of language. In doing so, he ultimately rejects any systematic and comprehensive account of language as developed in the *Tractatus*. Language is not a uniform phenomenon. Rather, the many uses of language must be examined individually; and from there one cannot generalize.

As a consequence, the meaning of an expression might vary from language-game to language-game. This may seem implausible at first. After all, it is not without reason that the same word can be used in different contexts. Wittgenstein attributes the initial plausibility of univocal meanings to superficial similarities in sentence structure. "Of course, what confuses us is the uniform appearance of words when we hear them in speech, or see them written or in print. For their *use* is not that obvious. Especially when we are doing philosophy!" (PI 2009 § 11; emphasis in original) So one must not be misled by the syntactic uniformity of sentences like "trees exist" and "numbers exist". The surface grammar of these sentences obscures the fact that "exist" is *used* in different ways. As Wittgenstein himself points out, the use of an expression is not always obvious. To shed some light on this issue, let us consider the following (non-exhaustive) list of uses of "exist".

1. After searching without success for a while, one discovers that there still is butter in the fridge.
2. When doing math homework, one comes to the conclusion that a (real-valued) solution to a given quadratic formula *exists*.
3. Given prior ignorance, one comes to learn that a biblical figure, like Jesus, actually *existed*.
4. The Higgs Boson, whose existence had been postulated in the 1960s, was confirmed to *exist* in 2012.

It can be seen that the concept of existence is used in a number of different ways and in relation to a number of different activities. Since its uses vary, so do its meanings. Of course, some meanings will be more similar than others. Thus, one must closely examine the language-games in which a specific existence claim is made in order to grasp the respective meaning.

4. Realist and Anti-Realist Language-Games

In this section, the Wittgensteinian strategy outlined above will be applied to the language-games of realists and anti-realists. By analyzing the practices they engage in, I hope to show that their individual language-games differ and that different meanings of existence are at play.

A first challenge to this undertaking arises from the observation that the concept of existence is rarely explicitly stated, since much of the analytic tradition agrees on the synonymy of existence and existential quantification. This assumption is encapsulated in the Quinean slogan "[t]o be is to be the value of a variable" (Quine 1948: 34). In light of Wittgenstein's earlier remarks on surface grammar, one should be particularly wary of this identification. While existence claims can be formalized as existentially quantified statements, this does not reveal their individual deep grammar. Arguably, this formalization causes even more confusion, since different expressions like "exist" and "there is" get substituted by the same existential quantifier. This suggests a semantic uniformity where there is none. The deep grammar of an existential claim only reveals itself when its use is taken into account.

Let us do this and look first at the realist's language-game. A realist will usually appeal to the trivial truth of statements like "there are prime numbers greater

than 100" and "there are green trees in the garden". From "there are prime numbers greater than 100" she will infer in a first step that there are prime numbers, and in a second step that there are numbers. Similarly, from "there are green trees in the garden" she will infer that there is a property that the trees in the garden have in common (*being green*), and from there that there are properties (cf. Thomasson 2015: 131). At this point, not much can be said about the uses of "exists". It will therefore be necessary to consider the contexts in which these existence claims are made.

Arguably, one does not start to question the existence of trees out of the blue. However, doubts about the existence of an empirical entity arise naturally when our perception is affected by external factors. For example, in dim light, we might wonder whether a figure outside the window is a tree or just a shadow, only to realize upon closer inspection that it is indeed a tree. Since we have convinced ourselves of the existence of at least one tree, we are justified in inferring that trees exist. Similarly, when confronted with the philosophical question of whether chairs exist, the ordinary person would point, albeit somewhat puzzled, to a nearby chair and infer that chairs exist. In this language-game, a specific witness for the entity in question (e.g., the tree outside the window, the chair over there) is given to assert its existence. Conversely, the absence of such a witness is taken as proof of non-existence. For example, we conclude that there are no green trees in the garden by observing for each tree that it is not green. More generally, we conclude that unicorns do not exist because we could not possibly point to a single witness in our (or anyone else's) experience.

Now suppose that the question is followed by the question whether numbers exist. This will usually cause confusion, since our previous use of "exist" does not extend to numbers in an obvious way. Of course, when pressed, the individual might admit that numbers do not exist, at least not in the sense of "exist" as used above. Since she cannot point to any number in her experience, the rules of the language-game suggest a negative answer to the existential question. However, anyone who understands the concept of numbers would not seriously question whether they can be encountered in the empirical world. Numbers are, after all, abstract entities, so it is reasonable to assume that the interlocutor has some other meaning of existence in mind. Thus, Wittgenstein considers "That depends on what you understand by '[exist]'" (PI

2009 § 47) as the appropriate response. "And that, of course, is not an answer to, but a rejection of, the question." (PI 2009 § 47)

We have seen that the introduction of numbers into an empirical context leads to confusion. So we will now turn to language-games for mathematical existence statements. Claims such as "there are prime numbers greater than 100" are commonly found in mathematics. There, existence claims can either be made by giving a witness (though we might not be able to refer to it, we just know that there is one) or by *reductio* arguments. In particular, the realist is licensed to infer the existence of numbers in the straightforward way sketched above. But unlike in the previous example, the mathematician does not need to know the value of a solution to know that it exists, i.e. she does not need to have an immediate witness available to her. It follows that these assertions of existence obey different non-logical rules than empirical assertions of existence. Again, some mathematical existence claims will be closer to empirical claims than others. I therefore do not exclude the possibility that there may be cases where the non-logical rules are fairly similar. In addition, mathematicians are only concerned with mathematical entities. A question about the existence of trees will not arise in such contexts. Thus, given the meaning of "exist" in the mathematical language-game, the mathematician will not be able to make sense of the question "Do prime numbers and chairs exist?" Again, the use of "exist" in a mathematical setting does not extend in an obvious way to empirical objects. This suggests that despite the similarities in surface grammar, the meaning of "exist" varies—even within the realist's setting. While she takes herself to be engaged in the same language-game, the Wittgensteinian reading reveals that she actually engages in multiple language-games.

But let us grant that the realist's comprehensive notion of existence captures (at least one possible) use of existence. Still, this is not how the anti-realist uses "exist"—even though both ascribe the identification of existence with existential quantification. When confronted with the realist's existential statement, "there is a property that the trees in the garden have in common," the anti-realist will appeal to the lack of logical clarity in ordinary language. Then she will resort to paraphrase, and take the paraphrased statement to reveal the underlying logical form of the claim. For example, she might paraphrase the above claim as "for every tree x in the garden, it holds that x is

green". In this way, it no longer seems as if properties are being quantified over. Whenever a paraphrase that only quantifies over concrete objects is available, the anti-realist will deny the existence claim of the unparaphrased statement. Whether an entity exists thus depends on whether quantification over it can be paraphrased away. This shows that the anti-realist uses *existence* as interchangeable with *necessarily being quantified over* (cf. Colyvan 2019). This, of course, is not a strategy endorsed by the realist, who may even deny the relevance of the availability of paraphrase for ontological discourse, and it is probably not what the ordinary person has in mind when making existential claims (cf. Thomasson 2015: 131). As a result, realists and anti-realists use "exist" in different ways. They disagree less about what exists than about what it means to exist.

It is only against the background of the same concept of existence that serious disagreement about the existence of a particular entity can arise. Although Quine rejects the existence of properties and propositions as semantic illusions, he admits the existence of numbers, which he considers indispensable for the truth of our mathematical theories. Conversely, Field (2016) adopts the Quinean notion of existence, but develops a strategy to make quantification over numbers redundant for the truth mathematical theories. On this basis, he concludes that numbers do not exist. Thus, it is Quine and Field who have a meaningful ontological disagreement, not Quine and the realist from above.

5. Consequences

In conclusion, it can be said that Wittgenstein advances a view reminiscent of ontological pluralism, since the meaning of existence differs from one language-game to another (cf. McDaniel 2017). Thus, each language-game establishes a different *way of being*. At the same time, Wittgenstein notes that the boundaries between different ways of being are fluid rather than fixed, since kinds themselves are dynamic (cf. PI 2009 §17). Depending on the criterion used to group entities into categories, one ends up with different kinds of entities. Entities that belong to different categories on one account could end up in the same category on another account. Thus, ways of being are not determinate, but relative to our method of ontological categorization.

Moreover, much of the realism/anti-realism debate turns into a verbal dispute on the Wittgensteinian account. This does not mean, however, that there is no way that realists and anti-realists could have a meaningful discourse. Similarly, Wittgenstein would not claim that there is no way that "exist" could be applied to both chairs and numbers. However, such a use of the concept of existence would require a corresponding language-game, the rules of which we have yet to learn. In particular, any confusion about the meaning of existence must be resolved at the outset. This shifts the focus from considerations of individual cases to the general meaning of existence. As a consequence, the risk of verbal disputes is limited. Moreover, it allows us to sharpen our understanding of existence and the questions it raises. It would lead to a fruitful discussion of the very nature of the entity in question if one finds that, despite using the same concept of existence, one still disagrees about the existence of some entity. This, in turn, will lead to a reevaluation of what it means to be a particular entity; and this is probably where the real core of the disagreement lies.

References

- Carnap, Rudolf (1950) "Empiricism, Semantics, and Ontology", *Revue Internationale de Philosophie* 4(11), 20–40.
- Colyvan, Mark (2019) "Indispensability Arguments in the Philosophy of Mathematics", *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta and Uri Nodelman (eds.) Spring 2019, Metaphysics Research Lab, Stanford University.
- Field, Hartry (2016) *Science Without Numbers - A Defense of Nominalism*, Oxford University Press.
- Hirsch, Eli (2002) "Quantifier Variance and Realism", *Philosophical Issues* 12, 51-73.
- Hirsch, Eli (2009) "Ontology and Alternative Languages", in: Chalmers, David; Manley, David; Wasserman, Ryan (eds.): *Metametaphysics. New Essays on the Foundations of Ontology*, Oxford University Press, 231-259.
- McDaniel, Kris (2017) *The Fragmentation of Being*, Oxford University Press.
- Quine, Willard V. O. (1948) "On What There Is", *The Review of Metaphysics* 2(5), 21–38.
- Thomasson, Amie L. (2015) *Ontology Made Easy*, Oxford University Press.
- Thomasson, Amie L. (2017) "Metaphysical Disputes and Metalinguistic Negotiation", *Analytic Philosophy* 58, 1-28.

Wittgenstein, Ludwig (2009) *Philosophical Investigations*, 4th ed. P. M. S. Hacker and Joachim Schulte (eds.), G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte (trans.).

Wittgenstein, Ludwig (2014) *Werkausgabe Band 1. Tractatus logico-philosophicus / Tagebücher 1914–1916 / Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp.

Reichenbach's Relativised A Priori

Luca Oliva (Houston, USA)

Abstract

Logical empiricists unanimously criticize the Kantian notion of *a priori* and related epistemic claims. Although cognitively relevant, the notion appeared untenable after the developments in geometry (Bolyai and Lobachevsky, Riemann, Minkowski, and Hilbert, among others) and physics (primarily, Lorentz and Einstein) revolutionized scientific research and its epistemic method. Nevertheless, a clear assessment of the Kantian *a priori* remains problematic because various readings revise or reject its properties, justifying Parrini's (2002) distinction between a *weak* and a *strong* rejection of Kant among logical empiricists. For instance, Schlick (1918), following Poincaré, argues for the conventional nature of the *a priori* by replacing the Kantian notion, based on sensible intuition, with that of implicit definition, primarily justified by Hilbert (1902). On the contrary, the early Reichenbach (1920) relativizes but retains the constitutive function of the *a priori*, dismissing its apodicticity alone. He argues for principles coordinating the theoretical and the empirical, which he sees as equivalent to Kant's synthetic *a priori* judgments. Notwithstanding Reichenbach's (1924, 1928) later acceptance of Schlick's criticism, his notion of relativized *a priori* aims to bridge the gap between abstract mathematical structures and concrete physical phenomena (Friedman, 2001; De Boer, 2010), maintaining the apriority's constitutive feature, also present in Carnap's *L-rules* (1928; Friedman, 1999). This paper shows Reichenbach's weak rejection of the Kantian *a priori* and its cognitive-metaphysical implications.

In light of Lorentz's and Einstein's new physics and its non-Euclidean geometry (Bolyai and Lobachevsky, Riemann, Minkowski, among others), logical empiricists unitedly reject Kant's version of *a priori*, especially its psychological and apodictic features (Friedman, 2007). Nevertheless, although transformed, its cognitive purpose is retained (Parrini, 1998). Instances of logical-empiricist variations of the *a priori* are the notions of (a) *relativized a priori* and related *coordinative definitions* (Reichenbach, 1920; Friedman, 2009), (b) *implicit definitions* (Schlick, 1918; Popper, 1959; Einstein, 1921; Giovannini-Schiemer, 2019), and (c) *L-rules* (Carnap, 1928). While (a) and (c), although in different ways (Friedman, 1999), maintain the constitutive character of the *a priori*, (b) only argues for its conventionality, following Poincaré (1902) and Hilbert (1899). These *weak* and *strong* rejections of the Kantian *a priori* (Parrini, 2022) betray its uneasy evolution through logical empiricism, showing irreducible inconsistency. However, if the *a priori* evolve according to empirical findings, in what sense is it still Kantian?

1. Constitution and apodicticity

Among the features of the Kantian *a priori* (see Anderson, 2014), logical empiricists isolate two. For them, *a priori* means “necessary and unrevisable, true for all time,” and /or “constitutive of the concept of the object of [scientific] knowledge” (Friedman 2001: 72). The distinction appears first in Kant, then occurs in Reichenbach (1920) and Carnap (1928). Kant’s *a priori* principles are absolutely (i.e., non-relatively) necessary or apodictically certain but also perform a “constitutive function with respect to *a posteriori* or empirical truths” (Friedman, 2001: 73). Such function makes the empirical cognition and confirmation of related truths possible. Therefore, *a priori* principles must be unrevisable by default.

The reason that *a priori* knowledge is in fact independent of empirical cognition or experience, for *Kant*, is that *a priori* knowledge yields the necessary conditions under which alone empirical cognition or experience can take place. Since they formulate the necessary conditions or rules for establishing empirical knowledge, *a priori* principles cannot themselves be similarly established; and it is in precisely this sense that they are prior to or independent of experience. (Friedman, 2001: 73)

However, logical empiricists challenge this type of unrevisability or independence from experience. Once they acknowledge that “those principles Kant took to be *a priori* can, after all, be revised,” says Friedman, “the way is then open, as it was for Reichenbach and Carnap, to retain Kant’s characteristic understanding of *a priori* principles as constitutive” (2001: 73) while dismissing the marks of necessity, unrevisability, and apodictic certainty (see Oliva, 2024).

Accordingly, for Reichenbach (1920), exact sciences rely on “the notion of a relativized yet still constitutive *a priori*” (Friedman, 2001: 71), although “*a priori* principles (both mathematical and physical) change and develop with the continual progress of empirical natural science” (Ibid), unlike Kant thought.

According to the traditional conception of the *a priori*, in which it means “justified independently of experience,” [...] any principle correctly characterized as *a priori* would perforce have to hold (if it does hold) entirely independently of all empirical findings and would thus have to hold “come what may”. (Friedman, 2001: 71)

Essential to this notion of relativized *a priori* (i.e., constitutive and revisable) is the relation between *a priori* and *a posteriori* cognition. What exactly does it mean for the *a priori* to represent (a) *necessary conditions* and (b) *constitutive principles* of empirical knowledge? Kant overlaps (a) and (b), holding that constitutive principles are necessary conditions (of the possibility of empirical laws). Still, Friedman warns us that (a) differs from a standard sense, “where A is a necessary condition of B simply if B implies A” (2001: 74). Indeed, in the Kantian sense, (a) entails (b), namely a normative framework for empirical knowledge.

To say that A is a constitutive condition of B rather means that A is a necessary condition, not simply of the truth of B, but of B's meaningfulness or possession of a truth value. It means [...] that A is a *presupposition* of B. (Friedman, 2001: 74)

Consider Newton's physics, where the law of universal gravitation uses the absolute acceleration concept, which doesn't have empirical meaning or application (within that physics) unless the laws of motion hold. So, “we know how to give empirical meaning and application to the law of universal gravitation,” explains Friedman, only “by presupposing that the laws of motion are true” (2001: 75). These laws work as *a priori* principles. So, if they are untrue, there cannot exist the frame of reference in which they hold, entailing that “the question of the empirical truth (or falsity) of the law of universal gravitation,” states Friedman, “cannot even arise” (Ibid).

2. Coordinative principles

For Parrini, the relativized *a priori* bridges the gap “between abstract mathematical structures and concrete physical phenomena” (Friedman, 2001: 78). This issue is central to the logical empiricist agenda. Schlick and Reichenbach identify a special class of non-empirical physical principles to solve it, which they call *coordinating or constitutive principles* and *conventions* (in Poincaré's sense), respectively. Van Fraassen exemplifies this class with the notion of measurements, which help determine the values of mathematical functions. Without them, a theory remains pure. It could never become empirical “if its terms were not linked to measurement procedures” (Van Fraassen, 2008: 115). This linkage raises the problem of coordination: “*How*

can an abstract entity, such as a mathematical space, represent something that is not abstract, something in nature?" (Van Fraassen, 2006: 537).

Reichenbach replies by pointing to modern physics (see Oliva 2024), where mathematical equations represent all processes. Still, the two sciences significantly differ. Indeed, "the truth of mathematical propositions depends upon internal relations among their terms," but "the truth of physical propositions," argues Reichenbach, "depends on relations to something external, on a connection with experience" (1965: 34). Consequently, we ascribe absolute certainty to (1) the former kind of assertions and probability to (2) the latter. But what about (3) their relationship? Let's see (1-3) in detail.

(1) Reichenbach endorses a quasi-structuralism in mathematics, where entities are determined by primitive definitions (i.e., axioms) whose terms rely on the other defining terms belonging to a shared framework.

The *mathematical object* of knowledge is uniquely determined by the axioms and definitions of mathematics. The definitions indicate how a term is related to previously defined terms. The mathematical object receives meaning and content within this framework of definitions through an analysis of its differences from and equivalences to other mathematical objects. (Reichenbach, 1965: 35).

Accordingly, the axioms present the mathematical rules for defining concepts (see Oliva, 2024). All concepts, including the fundamental ones (i.e., those occurring in the axioms themselves), are defined through relations. Reichenbach justifies his argument by referring to Hilbert. Consider his axiom of order II-3, stating that "Of any three points situated on a straight line, there is always one and only one which lies between the other two" (Hilbert, 1902: 4). Here, Hilbert describes the properties of 'point,' 'straight line,' and 'between' through a *non-exhaustive* definition, made complete solely by the totality of the axioms. All the entities involved (i.e., 'point,' 'straight line,' and 'between') have the axiom-stated properties, owing their nature to mutual relations that can change. For, in projective geometry, 'straight-line' and 'point' are interchanged, preserving the truth of related theorems since "their axiomatically defined relations are symmetrical for the two concepts" (Reichenbach, 1965: 35), although, as Schlick noticed (1974), our

intuition depicts the two concepts dissimilarly, ascribing different contents to the axioms.

(2) For Reichenbach, the method of representing physical events relies on mathematical equations. It defines one magnitude in terms of others by relating them to increasingly general magnitudes, up to the axioms. “Yet what is obtained,” he argues, “is just a system of mathematical relations,” ultimately lacking a statement of its significance, namely “the assertion that the system of equations is *true for reality*” (1965: 36).

The *physical object* cannot be determined by axioms and definitions. It is a thing of the real world, not an object of the logical world of mathematics. (Reichenbach 1965: 36)

(3) Hence, the internal coherence of mathematics doesn't suffice for physical truths, which further must entail a precise relation between equations and physical phenomena.

The physical relation can be conceived as a coordination: physical things are coordinated to equations. Not only the totality of real things is coordinated to the total system of equations, but *individual* things are coordinated to *individual* equations. The real must always be regarded as given by some perception. (Reichenbach, 1965: 36-7)

E.g., to name the earth a sphere, we must coordinate the geometrical spherical figure to a specific visual perception, i.e., a perceptual image of the earth, according to some primitive coordination principles (see Oliva, 2024). For this purpose, Reichenbach refers to Boyle's gas law, where we coordinate the formula $p \times V = R \times T$ to direct (e.g., feelings) and indirect (e.g., the position of a monometer's pointer) perceptions of gas. Indeed, “our sense organs mediate between concepts and reality” (Reichenbach, 1965: 37) – today's formula for the ideal gas is $PV = nRT$. Such coordination occurs in two moments. Accordingly, within the context of any particular scientific theory, Reichenbach identifies two types of cognitive principles: (a) the *axioms of connection* defining empirical laws that involve already well-defined terms and concepts, and (b) the *axioms of coordination*, namely non-empirical principles laid down antecedently to ensure that empirical well-definedness.

The peculiar nature of such coordination allows us to establish a correspondence between two sets by coordinating every element of one set with an element of the other. For this purpose, “the elements of each set must be defined,” argues Reichenbach, i.e., “for each element, there must exist another definition in addition to that which determines the coordination to the other set” (1965: 37). Yet, although the ‘equations’ (i.e., the conceptual side of the coordination) are uniquely defined, the ‘real’ (i.e., the side dealing with the cognition of reality) isn’t. Reichenbach overcomes this issue as follows.

The definition results from coordinating things to equations. Thus, we are faced with the strange fact that in the realm of cognition, two sets are coordinated, one of which not only attains its order through this coordination but whose elements are *defined by means of this coordination*. (Reichenbach 1965: 40)

Therefore, Reichenbach’s coordinating principles define reality according to our perceptions and mathematical equations. In this regard, they resemble Kant’s conceptualization of sensible intuitions, which applies rules on sensations, namely schemata and pure principles of understanding.

3. Synthetic a priori principles

Do Kant’s constitutively *a priori* schemata, mediating between algebraic abstractions and applied physics, anticipate Reichenbach's coordination?

Like Reichenbach’s, Kant’s actual object of cognition derives from the employment of formal structures on blind modifications of sensibility, namely raw sensations (see Oliva, 2024). Accordingly, the matter of the cognitive object represents a yet-to-be-determined empirical condition, *a posteriori*. Kant conceives of such determination in two steps, mainly focusing on its possibility or formality. (1) Categories unify the manifold intrinsic to our inner sense, shaping the corresponding *schemata*. Indeed, for the content of a category per se is initially derived from the logical structure of judgments alone, “it must be made applicable to objects whose form has thus far been specified solely by the pure forms of space and time,” say Guyer and Wood (1998: 10). So, schemata associate categories to a form or relation in intuition, particularly an inner temporal one. (2) The *principles of pure understanding* define the rules

for applying these schemata to empirical judgments based on our spatial outer sense. Accordingly, “the *use* of [those] schemata in turn depends upon judgments about the *spatial* properties and relations of at least some objects of empirical judgment” (Ibid).

The *analytic of principles* will accordingly be solely a canon for the *power of judgment* that teaches it to apply to appearances the concepts of the understanding, which contain the condition for rules *a priori*. (Kant 1787: A132/B171)

“Just like Kant’s synthetic *a priori* principles, principles of coordination assign conceptual structures to the realm of experience,” says De Boer, and “bridge the gap between the conceptual and the sensible” (2010: 517). For Reichenbach, “they ultimately define real objects and real events;” therefore, “we may call them constitutive principles of experience” (1965: 49). So, he refers to Kant’s schemata.

Therefore, Reichenbach’s early works evidence a weak rejection of the Kantian *a priori*. As De Boer argues, he “aimed to transform rather than abolish Kant’s notion of synthetic *a priori* principles” (2010: 508); for her, the differences with Kant have been overestimated by logical empiricists and their readers, such as Friedman.

Friedman weds Kant’s transcendentalism to Newton’s mechanics and Euclid’s geometry. If correct, Einstein’s new physics demands a detachment from the Kantian *a priori*, stemming from an obsolete paradigm. However, De Boer separates Kant’s synthetic *a priori* principles from Euclidian geometry and Newtonian physics, which instantiate but don’t demonstrate the validity of those principles. Instead, she thinks such validity has a metaphysical nature, traceable back to Leibniz and Hume. Friedman believes that Kant abandoned classic metaphysics (see 1992: 37-8), but De Boer contends that he seeks to reconcile metaphysics with Newton’s scientific paradigm. However, these two don’t overlap since the first can ground any science without restriction.

De Boer convincingly argues that the conditions constructing (*a*) an object of cognition and those building (*b*) a physical law differ. It’s always the case that (*a*) entails (*b*), but the opposite doesn’t hold.

[...] the synthetic *a priori* principles treated in the *Critique* merely constitute necessary rules for determining the spatio-temporal, law-governed relations between given representations - whatever the actual content of these relations may be. (De Boer 2010: 510)

“Unlike laws of physics [*b*], the principles of pure understanding do not depict the world, but constitute the ‘rules of the pure thinking of an object’ [*a*]” (De Boer, 2010: 510). These principles offer “perspectives that we must necessarily adopt,” says de Boer, “to turn phenomena into objects of knowledge” (Ibid). “Without such synthetic *a priori* root-principles, it would not be possible for us to establish laws of physics proper” (Ibid). So, De Boer sharply divides (*a*) and (*b*). Accordingly, “the synthetic *a priori* principles of pure understanding” represent “the root-principles by dint of which something can become an object of knowledge in the first place,” namely, they delimit the domain in which “something can be treated as an object” (Ibid). However, how does (*a*)’s grounding of (*b*) work?

Thus, far from telling us something about the world, the principle based on the pure concept of quantity merely states the rule that every intuited object has an extension and, hence, can be determined mathematically. According to Kant, it’s only on the basis of this principle that physics can apply pure mathematics to objects of experience (A165/B206). The category of substance, for its part, yields the rule that scientific knowledge must necessarily distinguish between that which changes over time and that which constitutes the self-identical substrate of such changes. Otherwise, scientists would neither be able to determine something as an object, nor to determine the relation between objects. (De Boer, 2010: 512).

De Boer’s analysis supports the weak rejection of the Kantian *a priori* by showing its similarity with Reichenbach’s coordinative principles. Although relativized, the *a priori* constitutive component must be retained as it grounds every cognitive object before being employed in a scientific theory.

4. Coordination and reality

Unlike Kant, Reichenbach holds that “the content of every perception is far too complex to serve as an element of coordination” (1965: 40). Before

coordination, we must sort out relevant from irrelevant aspects of our perception; namely, we must establish order among them. However, “such a coordination presupposes the equations, or the laws expressed in them” (Ibid). So, Reichenbach maintains that physical knowledge relies entirely on coordination. Indeed, “only a cognitive judgment,” he argues, that is an act of coordination, can decide whether the sensation of a tree corresponds to a real tree” (1965: 41), and not to a hallucination.

So, perceptions don't define what is real. Therefore, the elements of the universal set remain undefined since “one side of the cognitive process contains an undefined class” (1965: 42).

Thus, it happens that individual things and their order will be defined by physical laws. The coordination itself creates one of the sequences of elements to be coordinated. (Reichenbach 1965: 42)

Hence, coordination occurs between a given set of mathematical equations and a completely undetermined reality. Only assigning these equations to experience turns this latter into a proper domain of physics. Indeed, coordinating principles define the individual elements of reality and, in this sense, constitute the real object. Therefore, like Kant, Reichenbach holds that physics relies on rules unifying pure thought (exemplified by mathematics) and sensible experience.

Nevertheless, the two sides of knowledge maintain a mutual relationship. Indeed, “the defined side does not carry its justification within itself,” argues Reichenbach, as “its structure is determined from outside” (1965: 42). So, the coordination to undefined elements is restricted by experience, not arbitrary. Therefore, knowledge preserves an empirical determination or character.

We notice the strange fact that it is the defined side that determines the individual things of the undefined side, and that, vice versa, it is the undefined side that prescribes the order of the defined side. *The existence of reality is expressed in this mutuality of coordination.* (Reichenbach 1965: 42)

Therefore, this mutuality attests to what is real. It also guarantees truth, which consistently derives from correct coordination, correlating to experience data.

As Reichenbach states, “contradictions are discovered by observation” (1965: 43). To be true, a theory must continuously lead to consistent coordination. Like Schlick, Reichenbach consistently defines “*truth in terms of unique coordination*” (Ibid). So, perceptions play a crucial role in the cognitive process since they “*furnish the criterion for the uniqueness of the coordination,*” he claims (1965: 44). Similarly, About the correspondence of judgments with facts, Schlick states that, “a judgment that *uniquely designates* a set of facts is called *true*” (1974: 60).

Despite the differences, Reichenbach underlines the continuity with Kant. His theory of cognitive coordination straightforwardly answers Kant’s question, “How is pure natural science possible?” (B20). In a Kantian fashion, ‘possible’ has a logical, not a psycho-physical, meaning; “it pertains to the logical conditions of a coordination” (Reichenbach, 1965: 47). Hence, Reichenbach rephrases Kant’s question as follows, “*By means of which principles will a coordination of equations to physical reality become unique?*” (1965: 48). Accordingly, these epistemological principles of coordination “are equivalent to Kant’s synthetic a priori judgments” (Ibid).

5. Conclusion

Hence, revisability and sensible dependency define Reichenbach’s ‘weak rejection’ in 1920. Later (1924, 1928, 1936), he softened, without ever endorsing it, though, his early criticism of Schlick’s ‘strong rejection’ of Kant’s *a priori*, based on the conventionalist characterization of this latter and the denial of its constitutive function. The idea that constitutive scientific principles can be *a priori* and revisable at the same time perfectly instantiates Parrini’s notion of ‘weak rejection,’ balancing continuity and discontinuity with Kant.

References

- Carnap, Rudolf (2005 [1928]) *The Logical Structure of the World*, Chicago and La Salle: Open Court.
- De Boer, Karin (2010) “Kant, Reichenbach, and the Fate of A Priori Principles”, *European Journal of Philosophy* 19/4, 507-31.
- Friedman, Michael (1992) *Kant and the Exact Sciences*, Cambridge, MA: Harvard University Press.

Friedman, Michael (1999) *Reconsidering Logical Positivism*, New York: Cambridge University Press.

Friedman, Michael (2001) *Dynamics of Reason*, Stanford, CA: CSLI Publications.

Friedman, Michael (2007) "Coordination, Constitution, and Convention: The Evolution of the A Priori in Logical Empiricism", in: A. Richardson and Uebel (eds), *The Cambridge Companion to Logical Empiricism*, New York: Cambridge University Press, 91-116.

Friedman, Michael (2009) "Einstein, Kant, and the Relativized A Priori", in: M. Bitbol, P. Kerszberg, J. Petitot (eds.), *Constituting Objectivity*. Springer Science, 35-50.

Guyer, Paul and Woos, Allen (1998) *Introduction to the Critique of Pure Reason*, in: Immanuel Kant, *Critique of Pure Reason*. P. Guyer/A.W. Wood (eds). Cambridge: Cambridge University Press, 1-80.

Hilbert, David (1902 [1899]) *The Foundations of Geometry*, E. J. Townsend (transl.). Chicago: Open Court 1902. First published in German as *Grundlagen der Geometrie*. Leipzig: Teubner, 1899.

Kant, Immanuel (1998 [1787]) *Critique of Pure Reason*. P. Guyer/A.W. Wood (eds). Cambridge: Cambridge University Press.

Oliva, Luca (2024) "Parrini on the A Priori in Logical Empiricism", *Humana Mente*, 2024: forthcoming.

Parrini, Paolo (1998) *Knowledge and Reality. An Essay in Positive Philosophy*, Springer. Modified version of Parrini, Paolo (1995), *Conoscenza e Realtà*. Roma-Bari: Laterza.

Parrini, Paolo (2002) *L'Empirismo Logico - Apetti Storici e Prospettive Teoriche*, Roma: Carocci.

Poincaré, Henri (1905 [1902]) *Science and Hypotheses*, New York: The Science Press.

Reichenbach, Hans (1936) "L'empirisme logistique et la désagrégation de l'a priori", *Actes du Congrès international de philosophie scientifique*, Paris 1935, Vol. I: *Philosophie scientifique et empirisme logique*, Paris: Hermann, 28-35.

Reichenbach, Hans (1965 [1920]) *The Theory of Relativity and A Priori Knowledge*, Berkeley and Los Angeles: University of California Press. Originally published as *Relativitätstheorie und Erkenntnis Apriori*. Berlin: Springer, 1920.

Reichenbach, Hans 1969 [1924] *Axiomatization of the Theory of Relativity*, M. Reichenbach (transl.), Los Angeles: University of California Press. Originally published as *Axiomatik der relativistischen Raum-Zeit-Lehre*, Braunschweig: Vieweg, 1924.

Reichenbach, Hans 1958 [1928] *The Philosophy of Space and Time*, M. Reichenbach and J. Freund (transl.). New York: Dover. Originally published as *Philosophie der Raum-Zeit-Lehre*, Berlin: de Gruyter, 1928.

Schlick, Moritz (1974 [1918]) *General Theory of Knowledge*, New York-Wien: Springer. First published in German as *Allgemeine Erkenntnislehre*, Berlin: Springer 1918.

Van Fraassen, Bas C. (2008) *The Problem of Coordination*. Ch. 5 in *Scientific Representation: Paradoxes of Perspective*, New York: Oxford University Press, 115-39.

Real Responses vs. Judgments

C. Naomi Osorio-Kupferblum (Vienna, Austria)

Abstract

Response-dependent (R-D) properties have a big epistemological advantage: when we are the responders, they give us real knowledge of what their bearers can do or cause. But accounts vary substantially with respect to the underlying metaphysics, and the epistemological advantage is easily lost. In this paper, I explain how this occurs in Pettit's influential account.

I begin by outlining the epistemological motivation for dealing with R-D properties, in particular for some, more demanding, empiricist theories of knowledge. I then explain how dispositional accounts of R-D properties, like Johnston's, invite in accounts involving judgments, like Pettit's. In Pettit's account, responses are effectively judgments of salient similarities between objects, and thereby second-order, so that it is not the properties, but our concepts of the properties that are "response-privileging" and thereby ultimately R-D. This account is then extended to all concepts. Pettit thus gives us a R-D genealogy of concepts, but relinquishes the epistemological asset inherent in (first-order) accounts of R-D properties as consisting in ("real") response events.

1. Introduction

Response-dependent (R-D) properties, when first discussed under that name in the late 1980s, were of interest because they seemed apt to straddle the divide between subjectivity and reality for some properties. Modelled on Locke's (and probably Galilei's) secondary qualities – prominently exemplified by colours – R-D properties were considered those we ascribe to a thing when it elicits a certain kind of response in other objects, usually humans. So, if moral properties, for instance, were R-D, they would be both intimately linked to our responses and somehow really "out there". But how to define 'response'? Some (prominently Philip Pettit and Crispin Wright) seemed to think judgment could count as a response, and from judgment it is a small step to 'global response-dependence' (Pettit 1991), the idea that ultimately all properties are tied to concepts based on judgments and concepts therefore response-dependent. Pettit's view contains aspects important for empiricism. However, it takes us to an idealist version, so from a realist perspective, the account goes astray in a way that is highly detrimental to knowledge.

In this paper, I will discuss the epistemological disadvantage entailed by Pettit's view. I will start by outlining the epistemological motivation for dealing

with R-D in the first place. I will then show where judgment slips in, and finally use the epistemological aspect to explain why metaphysically neither judgments are responses, nor concepts response-dependent.

2. The epistemological motivation

Even though the question what are, and what are not R-D properties, is metaphysical, the interest in R-D properties surely arises from a quest for knowledge. We need to understand that quest to be able to get a grip on R-D properties. So let us start with examples concerning knowledge, including a question Wittgenstein raised.

1.a. Coming home, I notice that the door is not locked. So, my daughter is probably at home. Why think that? Well, as I locked it when I left, someone must have come and unlocked it; of the people who have a key, by far the most likely person to have done so this afternoon is my daughter. So, I have good reason to suppose that she has come home. Moreover, unless she forgot to lock the door when she left (which she doesn't tend to do), she must still be here.

1.b. I open the door and see my daughter's bag and shoes in the entrance. So, I am now convinced that she is at home.

1.c. I say "Hello, darling!" and she replies "Hi, mom!" from her room. I now know that she is at home.

At 1.a. I have good reason to guess that my daughter is at home – the likelihood is very high. At 1.b., I abduct from the evidence that she is at home; it is still possible that she might have left wearing different shoes and without that bag, and that she forgot to lock the door when leaving, but this is highly unlikely. So at 1.b. I am justified in supposing her to be at home – so much so that my conviction prompts me to shout a greeting. Had she not replied, I would have looked around the flat to see where she was and what she was doing that kept her from hearing, or answering, my greeting. Nevertheless, at 1.b., my taking her to be at home is an inference to the best (one might say, the obvious) explanation; however, it does not quite amount to knowledge.

At 1.c., by contrast, I am in direct epistemic contact with her – she is talking to me; so, I now *know* that she is at home. This is a classic case of the knowledge “by acquaintance” experience affords us. It would take a blatantly absurd construal to raise doubt about it – something along the lines of her voice coming from a loudspeaker, herself hologrammed into her room when I go to ask her how her day had been, etc. but nothing short of deep brain manipulation can defeat a mother’s kiss. (Two less polemic ways to reject such sceptical fantasies are Ayers’ (2019) shifting of the burden of proof to the sceptic, and Williamson’s (2002) account of taking knowledge logically first.)

Now, compare this to Wittgenstein’s remarks about pain (PI §246 and §288). He points out that it doesn’t make sense to say “I know that I am in pain” because doubting it would make no sense – there is no way to be mistaken about this. Pain is a WYSIWYG (“what you see is what you get”) feeling, as it were, – it is just what one feels and there is no other direct epistemic access to it than to feel it. When you feel it, you are in pain; when you don’t feel it, you’re not. (Note the contrast with biologically important information pain conveys by being located in a specific part of our body and about its possible cause by its quality. We can be mistaken about both: pain can be referred, and quality can mislead, as when cold is mistaken for heat, for instance. Note also the contrast with another person’s pain; *pace* Wittgenstein, surely our epistemic access to another person’s pain is as in 1.b., not 1.c.)

So, one might think that the subject who is in pain has perfect knowledge of their being in pain – except, if knowing something implies that one might be ignorant about that fact. If (human) knowledge requires the possibility of ignorance, one’s own pain is not knowable just because of our infallibility about. But without that logical restriction, this infallibility would make pain the empiricist’s ideal case of knowledge by experience. What a pity that we can’t have that level of certainty in our knowledge of other things! But: can’t we really?

There is one sort of epistemic relation to other things that is halfway between example 1.c and pain, and it is the sensations and feelings other things can cause in us. Looking at this ball, I see that it is red, I feel that it is soft and that its surface is smooth. These sensations I have are as first-hand and infallible as pain, but they are of another object which causes them. Nevertheless, my

sensations are just what these properties consist in. So, the ball's softness comes about in its interaction with me (my fingers). Likewise, its redness comes about in its interaction with me (my visual organs). Long before we even get to concepts of, and names for, those properties, if ever we do, there are the sensations caused in us – sensations that are as WYSIWYG and independent of concepts as pain.

However, as these sensations arise in interaction, they are subject to much variability. Different objects can produce like sensations; the same object produces different sensations, even of the same type, depending on variation in what I myself contribute to the interaction and to variation in the circumstances of our interaction. Moreover, the ball's surface feels smooth *to me* – it probably doesn't to an ant; and it is soft to the touch of my fingers, not that of a feather, softness being one of those covert comparatives. But despite their variability and thereby, in one respect, unreliability, the effects produced in me, i.e. the ways I respond to the other object, are as immediately present to me as pain; they consist in just that sensation, whatever it is like at that time, so in that sense, they are as infallible as pain. They are also as real as pain or, in fact, as the breaking of glass under the impact of a stone, as they arise from direct contact with another object. So, they combine the advantages of pain and direct perception as in 1.c, long before any sort of inference or judgment take place.

So, what sort of knowledge do we gain from our sensible responses to other things? Not much with respect to what the other things are like 'in and of themselves', or what about those things causes the responses, but a lot with respect to what they can do or cause. And that information is as certain as pain when we are the responder. It is this certainty that recommends these properties as primary building blocks in the epistemologies of philosophers more empiricist than Williamson and less defiant than Ayers. Particularly those of the sincerity of Carnap and the modesty of Locke will be drawn to them. In Carnap's philosophy, they appear as 'elementary experiences' at the ground level of his *Aufbau*. In Locke's, they are the simple ideas we obtain of secondary qualities, and on which all our knowledge depends. (Interpreting his secondary qualities as an 'error theory' seems to me to put the emphasis in entirely the wrong place; Johnston (1998) shows why.) Both are nominalists and derive concepts of properties from these experiential atoms by some form

of abstraction. This is not the place to discuss them further; suffice it to say that they started from infallibly given responses in order to ascribe R-D properties to the objects that caused them because they are real, given, and – for Locke (I think), but sadly not Carnap – because they therefore afford us real knowledge of their causes.

3. Enter judgment

Nominalists can build R-D properties from the atomic responses an object elicits. But property realists must give their explanation another direction. To them, responses are either caused by tropes or by instances of (multiply instantiable) properties. But as responses are not usually permanent, and moreover, vary, objects either only have the relevant properties temporarily, or they are not permanently manifest. Therefore Johnston, for instance, considers them dispositions. The appeal of dispositional accounts is probably that they seem more in line with our conceptual practice of ascribing the property to the object that causes the response. Thus, we call the ball red, soft and smooth, and think and speak of these properties as on a par with the ball's intrinsic properties. Reducing the responses away is thus pragmatically well motivated. The trouble is that we thereby lose sight of the one thing that guaranteed our knowledge of them: our responses. On most response-dispositional accounts, these properties need no longer be manifest for an object to have them. But then, they are on a par with genuine dispositions, like fragility, in that we may have no direct epistemic access to them but only infer from other – perceivable – properties that the object has, or must have, that disposition.

This is what leads to judgment, and in fact, the R-D debate has mostly not been about metaphysics, but about property *concepts*. And concepts are, of course, needed for making judgments. The issue became how to define, and further, when to correctly apply, property concepts involving judgments. Thus, Wright's account is for 'judgment-dependent' properties.

Pettit's approach is more astonishing. He gives a description of object and response, much of which is congenial to empiricists (hence this discussion of it), but then he inserts a judgment between response and property. The reason is that he, too, is concerned only with the concept; interestingly, he calls the

concept rather than the property “response-dependent” or, initially, “response-prioritising”; we will see why in the next section where I outline his view. Building his account on salient similarities, he ends up extending his theory to all property concepts – global response-dependence – at the expense of R-D property realism and having changed the metaphysics.

4. Losing the reality of response-dependent properties

We have seen how dispositional views give a metaphysically mutilated account of R-D properties, inviting in others that involve concepts and judgment. One example is Johnston’s view that secondary-quality concepts should be represented as response-dispositional; for properties like smoothness or redness, these responses are sensations. The property is then to be conceived as the disposition which is manifested to *normal* perceivers under *normal* conditions by that sensation. (Johnston 1989 as summed up by Pettit 1991, my emphasis) Pettit notes that

The sensations are not judgments but they lead observers to make judgments. (Pettit 1991:597)

And it is the judgments Pettit is interested in. He follows Johnston in distinguishing response-dispositional from response-privileging property concepts. For both it is a priori that x is F iff x looks F to normal observers under normal conditions, so that observers can neither be ignorant, nor in error concerning x ’s F -ness. However, response-dispositional concepts like ‘red’ acquaint us with the nature of the property, whereas response-privileging concepts, such as ‘water’, don’t because the response is now to salient similarities, i.e. the relationship *between* bearers of the property, not the bearers themselves. Nevertheless, assuming that we have learnt what water looks like under normal conditions, ignorance and error are still ruled out.

Pettit then extends this account of Johnston’s to all judgments based on some salient similarity (Pettit 1991:598), and in due course across the board to all our concepts. The idea is that rather than responses to the object, it is the salient similarities to exemplars, or between objects, that form the basis of our judgments. However, there are, of course, deviant cases when classical exemplars don’t display the right sort of property. Pettit explains that we learn

to find a feature that allows us to count out those cases. Moreover, we assume that both differences in our own judgments in like cases, and inter-personal differences of judgment are due to such features. We thus develop a practice of self-correction in non-normal situations.

This account gives us a genealogy of concepts: an ‘ethocentric’ “account of the conditions of response and practice under which [a concept] emerges and becomes accessible.” (Pettit 1991:601) Although Pettit stresses that his account does not focus on application, but on possession conditions (1991:604), possessing the concept means that we can apply it in a way that makes us competent participants in a discourse, discourse that is realist in the sense that it takes statements to be descriptive of objects existing independently of that discourse. Pettit thinks that this is enough realism. He thinks he doesn’t need the more demanding “cosmocentric” realism, a view of reality as something we can always be mistaken about because there is no guarantee that we are able to make suitable contact with the objects of discovery. (Pettit 1991:590)

However, what his account actually does, is to explain how our language practice latches on to experience. It tells us how we acquire concepts rather than knowledge of the world. It is indicative that Pettit has to take our disposition to find something similar in the right sense to be “sure-fire” (1991:599, fn21) – he has dropped the genuinely sure-fire knowledge inherent in our responses and so has no other way of safeguarding the impossibility of ignorance and error on which his account builds. That move may be fine for language acquisition, where large-scale consistency in personal and interpersonal use of words is sufficient for communication, but it is disastrous for scientific investigation. In putting responses in the cases of red and smooth in the same pot as the salient similarity of different portions of water, he treats recognition just like experience. In ignoring the reality of the event of red or smooth sensation, he relinquishes the most certain knowledge we have of the world and leaves us with ordinary knowledge by acquaintance – and this, moreover, not of the world but of the correct application of concepts.

But concepts are at the root of the problem. Judgments require criteria against which the judgment is made. Such criteria are conceptual in nature, so judgments require concepts. But secondary qualities don’t – the response they

depend on is an event for which the responder needs no concept whatever of the property their response grounds. So, Brynjarsdóttir (2008) is right that the criteria for the response-dependence of properties are different from, and in fact independent of, those of concepts; they concern their ontology and what having the property consists in. The property's dependence on the response is ontological, not logical, nor much less a matter of practice or convention. So R-D properties are not concepts and responses are not, or not usually, judgments.

This applies not only to the sort of judgment envisaged by Pettit, but also to Wright's. Although Wright distinguishes along Lockean lines between properties we merely *detect* in our responses (these would be Locke's primary qualities, e.g. squareness) and those *projected* by the responses, the responses are judgments and the projected ones include "judgements of colour or shape, or moral judgements, or mathematical judgements, for instance" (Wright 1992:108) (I take these to include social or 'conferred' properties; cf. Ásta (2008), Passinsky (2020). Paseau 2012 shows that mathematical properties do not in fact meet Wright's criteria for judgment-dependence.) So, Wright too, loses sight of the ontological aspect of response-dependence and, like Pettit, relinquishes the epistemological advantage inherent in the reality of responses.

5. Conclusion

What distinguishes R-D properties is that they come about in interaction between the object to which we ascribe the property and another object from which the first one elicits a response. Responses are events; although they take place – under certain circumstances – because the objects involved have the non-R-D properties they have, the R-D property is only correctly ascribed to objects if they elicit the response, i.e. cause the effect the property is about. So, explanations that prescind the response give us the cause without the effect. (Dispositional accounts moreover withdraw to modality where instead reality is available.) But accounts, like Pettit's, that shift the cause to a second-order response, making the property judgment-dependent, likewise prescind the – first-order – event. They have thereby changed the property's metaphysics, and undermine both the property's reality and its epistemological advantage over other properties.

References

- Ásta (2008) (published under "Ásta Kristjana Sveinsdóttir") "Siding with Euthyphro: Response-Dependence and Conferred Properties", *European Journal of Philosophy* 18, 1, 108-125.
- Ayers, Michael (2019) *Knowing and Seeing*, Oxford: Oxford University Press.
- Brynjarsdóttir, Eyja (2008) "Response-Dependence of Concepts Is Not for Properties", *American Philosophical Quarterly* Vol. 45, 4, 377-386.
- Carnap, Rudolf (1928) *Der logische Aufbau der Welt*, Berlin: Weltkreis.
- Johnston, Mark (1989) "Dispositional Theories of Value", *Proceedings of the Aristotelian Society, Supplementary Volume* 63, 139-74.
- Johnston, Mark (1998) "Are Manifest Qualities Response-dependent?", *The Monist*, 81,1, 3-43.
- Locke, John (1689/1979) *Essay Concerning Human Understanding*, Oxford: Clarendon.
- Paseau, Alexander (2012) "Against the Judgment-Dependence of Mathematics and Logic", *Erkenntnis* 76, 23–40.
- Passinsky, Asya (2020) "Social Objects, Response-dependence, and Realism", *Journal of the American Philosophical Association*, 6,4, 431-443.
- Pettit, Philip (1991) "Realism and Response-Dependence", *Mind* 100, 4, 587-626.
- Williamson, Timothy (2002) *Knowledge and its Limits*, Oxford: Oxford University Press.
- Wittgenstein, Ludwig (1960) *Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp.
- Wright, Crispin (1992) *Truth and objectivity*. Cambridge, MA: Harvard University Press.

Acknowledgement

This paper presents a very small part of work completed during the FWF-funded project T-1103, and builds in particular on research carried out at Oxford in spring 2023. During that stay I benefitted greatly from conversations with several people, but special thanks are due to Anita Avramides, Alex Paseau, and most of all, Michael Ayers.

Vergleichende Analyse der *Lebensform* und *Lebenswelt* bei Wittgenstein und Habermas

Christian Pflügl (Vienna, Austria)

Abstract

Das Sprachspiel bei Ludwig Wittgenstein und der Begriff der *Lebenswelt* bei Jürgen Habermas sind zentrale Konzepte in der Philosophie der Sprache, die jeweils den Fokus auf die Praxis des Sprechens und die intersubjektive Bedeutungskonstitution legen. Wittgenstein stellt in seinen *Philosophischen Untersuchungen* die These auf, dass die Bedeutung eines Wortes durch seinen Gebrauch in der Sprache und dem jeweiligen Sprachspiel bestimmt wird, was einen Bruch mit der traditionellen Vorstellung von Bedeutung als Repräsentation darstellt. Er betont, dass Regelfolgen eine öffentliche Praxis ist und die Interpretation von Regeln innerhalb einer Sprachgemeinschaft erfolgt, wodurch Kommunikation erst möglich wird. Die Voraussetzung für das Regelfolgen stellt der Begriff der *Lebensform* dar. Dieser Begriff der Lebensform soll mit dem Begriff der Lebenswelt aus der *Theorie des kommunikativen Handelns* von Habermas kontrastiert werden. Die Lebenswelt besteht aus kulturellen Überzeugungen, sozialen Praktiken und individuellen Fähigkeiten, die es ermöglichen, über die Inhalte von sozialen, objektiven und subjektiven Welten zu kommunizieren und diese auszuhandeln. In dieser Arbeit sollen die Parallelen und Unterschiede zwischen Wittgensteins Lebensform und Habermas' Lebenswelt in ihrer Rolle als konstitutives Element der Sprachpraxis und deren Genese aus der sozialen Praxis heraus analysiert werden.

Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache. (PU 2009: §43)

Diese auch über die Grenzen der Philosophie hinaus bekannte Definition gibt Wittgenstein in seinem posthum veröffentlichten Spätwerk, den *Philosophischen Untersuchungen*, auf die Fragen nach der Bedeutung von Wörtern an. Somit gilt, dass nach Wittgenstein die Bedeutung eines Wortes durch seinen Gebrauch in der Sprache in seinem jeweiligen Sprachspiel determiniert ist. Hierbei handelt es sich um einen Bruch mit der bis dahin allgemein vertretenen Auffassung, dass die Bedeutung eines Wortes sein Repräsentationsinhalt ist, also etwas repräsentieren muss, um etwas bedeuten zu können. Man könnte wohl aber auch argumentieren, dass dies ebenfalls auf die Wittgenstein'sche Definition zutrifft. Worte repräsentieren bei Wittgenstein anstelle von Ideen oder Dingen ihren Gebrauch in einer Sprecher*innengemeinschaft.

Um das Sprachverständnis Wittgensteins verstehen zu können, ist es notwendig, die Praxis des Sprechens zu verstehen. Wie sind also die Regeln

gefasst, die den Gebrauch von Wörtern für Sprecher formulieren? Gibt es diese Regeln überhaupt? Um uns nun dem praktischen Verständnis von Sprache bei Wittgenstein zu nähern, werden wir das Regelfolgen bei Wittgenstein behandeln. In den Paragraphen §§139–242 der *Philosophischen Untersuchungen* geht es um das Regelfolgen und das spätere Regelfolgen-Paradoxon. Hier versucht Wittgenstein zu widerlegen, dass es sich beim Regelfolgen um einen mentalen oder kausalen Zustand handelt. Das Wittgensteinsche Verständnis wird am deutlichsten in Paragraph §202: „Darum ist ‘der Regel folgen’ eine Praxis. Und der Regel zu folgen *glauben* ist nicht: der Regel folgen. Und darum kann man nicht der Regel ‘privatim’ folgen, weil sonst der Regel zu folgen glauben dasselbe wäre, wie der Regel folgen.“ (PU 2009: §202)

Hier wird deutlich, dass einerseits Regelfolgen kein mentaler Prozess sein kann, da es ansonsten möglich wäre, Regeln privat zu folgen. Auf der anderen Seite wird hier deutlich, dass Regelfolgen einer Öffentlichkeit bedarf. Bezüglich des Regelfolgen-Paradoxons bei Wittgenstein gibt es weitläufige Debatten innerhalb der philosophischen Gemeinschaft. Es ist nicht klar, wie die Passagen der *Philosophischen Untersuchungen* zu deuten sind. Es gibt zwei Strömungen der Auslegung des Regelfolgen-Paradoxes. Die eine ist die „straight interpretation“, die von John McDowell vertreten wird. Diese geht davon aus, dass das Paradoxon des Regelfolgens von Wittgenstein zwar angeführt, aber von ihm abgelehnt wird. Das „sceptical argument“ von Saul Kripke geht davon aus, dass das Paradox des Regelfolgens bei Wittgenstein ernst genommen werden muss. Daraus folgt, dass - was Wittgenstein in §202 zum Ausdruck bringt - es kein privates Regelfolgen gibt. Ob ein Sprecher also einer Regel gefolgt ist oder nicht, hängt nur davon ab, ob die Sprechergemeinschaft dies als einen regelkonformen Sprechakt akzeptiert oder nicht. Man kann nur dann einer Regel folgen, wenn die Gemeinschaft dies als Folgen der Regel interpretiert. Regeln werden somit durch die Praxis des Regelfolgens in einer Gemeinschaft konstituiert und erzeugt. In dieser Arbeit werden wir die skeptische Auslegung des Regelfolgens als geltende Grundlage benutzen.

1. Der Begriff der Lebenswelt bei Habermas

Wir wollen nun die Sprachphilosophie Wittgensteins und vor allem die Begriffe der Lebensform und der Praxis des Sprachspiels mit der *Theorie des kommunikativen Handelns* von Habermas und hier im Speziellen seinen Begriff der Lebenswelt kontrastieren. Was genau ist nun die Lebenswelt bei Habermas? Bei Husserl, auf den die phänomenologische Neuinterpretation des Begriffs der Lebenswelt zurückgeht, meint der Begriff der Lebenswelt einerseits die Wahrnehmung und Rechtfertigung des Subjekts von sich selbst im Widerspruch zur Welt und die realisierte, praktische Lebenswelt, in der das Leben eines Subjektes stattfindet, basierend auf den Gemeinschaften und Institutionen denen es angehört und nicht angehört (vgl. Beyer 2020).

Im Rahmen seiner Kommunikationstheorie reformiert Habermas den Begriff der Lebenswelt und führt ihn als ein intersubjektives Konzept, das die Grundlage jeglicher Kommunikation darstellt, in seine Kommunikationstheorie ein. In seiner Kommunikationstheorie unterscheidet Habermas drei Welten, welche „verständigungs-orientiert handelnde Subjekte ihren gemeinsamen Situationsdefinitionen zugrunde legen“ (Habermas 2011: 183). Diese drei Welten sind Akteur-Welt-Beziehungen eines Subjekts zu der Welt, in der es lebt. Die Welten werden von Habermas wie folgt definiert:

- zu etwas in der objektiven Welt (als der Gesamtheit der Entitäten, über die wahre Aussagen möglich sind); oder
- zu etwas in der sozialen Welt (als der Gesamtheit legitim geregelter interpersonaler Beziehungen); oder
- zu etwas in der subjektiven Welt (als der Gesamtheit der privilegiert zugänglichen Erlebnisse, die der Sprecher vor einem Publikum wahrhaftig äußern kann) (Habermas 2011: 183-184)

Diese drei Welten speisen den Inhalt kommunikativen Handelns und sind es, worauf sich Subjekte beziehen, wenn sie kommunizieren. Diese Welten sind aber nicht getrennt voneinander, da es in der sprachlichen Praxis kaum möglich ist, auf nur eine dieser Welten zu referenzieren. Das heißt, jede Aussage referenziert immer auf alle drei Welten, auch wenn in den Äußerungen nur eine der drei referenzierten Inhalte hervorgehoben wird. Verständigung ist bei Habermas die Einigung der Kommunikationsteilnehmer über die Gültigkeit des Gesagten (vgl. Habermas 2011: 183),

Einverständnis die intersubjektive Anerkennung des Geltungsanspruchs, den der Sprecher für sie erhebt. Selbst wenn eine Äußerung nur einem Kommunikationsmodus eindeutig zugehört und einen entsprechenden Geltungsanspruch scharf thematisiert, stehen die Kommunikationsmodi und die ihnen entsprechenden Geltungsansprüche untereinander in einem intakten Verweisungszusammenhang. (Habermas 2011: 184)

Der Begriff der Lebenswelt ist etwas komplexer. Um ihn zu attribuieren, scheint es sinnvoll, sich zuerst dem Begriff des Horizonts zu widmen. Unter dem Horizont versteht Habermas den Ausschnitt der Lebenswelt, der durch eine präsente Situation aktualisiert wird. Dieser Abschnitt wird bedingt durch den Handlungsspielraum, den eine gegebene Situation setzt, und dadurch die Kommunikationsmöglichkeiten vorgibt. Habermas beschreibt die Handlungssituation wie folgt: „Die Handlungssituation bildet für die Beteiligten jeweils das Zentrum ihrer Lebenswelt; sie hat einen beweglichen Horizont, weil sie auf die Komplexität der Lebenswelt verweist.“ (Habermas 2011: 188). Erst der Horizont ermöglicht eine Kommunikation, da er durch die realisierte Handlungssituation die Möglichkeit des Sprechens über Sachverhalte ermöglicht. Das Gleiche gilt nun auch für die Lebenswelt, die immer als eine Art Hintergrund vorhanden ist, aber für die Subjekte nur durch den Horizont wahrnehmbar wird (vgl. Habermas 2011: 188).

Habermas beschreibt die Lebenswelt nun wie folgt:

[...] die Lebenswelt durch einen kulturell überlieferten und sprachlich organisierten Vorrat an Deutungsmustern repräsentiert denken. Dann braucht die Rede von einem Verweisungszusammenhang, der die Situationsbestandteile untereinander und die Situation mit der Lebenswelt verbindet, nicht mehr im Rahmen einer Phänomenologie und Psychologie der Wahrnehmung erklärt zu werden. Verweisungszusammenhänge lassen sich vielmehr als Bedeutungszusammenhänge begreifen, die zwischen einer gegebenen kommunikativen Äußerung, dem unmittelbaren Kontext und ihrem konnotativen Bedeutungshorizont bestehen. Verweisungszusammenhänge gehen auf grammatisch geregelte Beziehungen zwischen Elementen eines sprachlich organisierten Wissensvorrats zurück. (Habermas 2011: 188-189)

Die Lebenswelt ist also eine Menge an geteilten Überzeugungsätzen, die es Sprecher*innen erlaubt, die Inhalte von sozialen, objektiven und subjektiven Welten untereinander auszuverhandeln. Diese Welten können nicht ohne einen gemeinsamen, geteilten, nicht hinterfragbaren Bedeutungsrahmen bestehen, welchen die Lebenswelt für sie darstellt. Die Beziehung zwischen Lebenswelt und den drei Welten ist reziproker Natur, da die drei Welten durch die Schablone der Lebenswelt bereits inhaltlich vordeterminiert sind. Nur so kann ein intersubjektiver Prozess des Ausverhandelns über ihre Inhalte vollführt werden. Dafür braucht es die Lebenswelt als Hintergrund, als ein Metaregulativ, das es den Sprecher*innen ermöglicht, Dinge ähnlich genug zu sehen, damit sie sich darüber austauschen können. Die Lebenswelt ist der Hintergrund, der Intersubjektivität erst ermöglicht.

2. Sprachspiel und Lebenswelt

Beginnen wir nun damit, das Paradox des Regelfolgens bei Wittgenstein weiter zu illuminieren. Regelfolgen ist eine Praxis, die der Öffentlichkeit bedarf, somit nicht alleine vor sich selbst möglich ist. Nun eröffnet sich aber die Frage, wie Kommunikation überhaupt möglich ist, wenn es unmöglich sein soll, Regeln zu folgen. Wieso spricht nicht jede Person eine eigene Sprache? Wie ist es möglich, sich ohne Regeln zu erklären, wie verständliches Sprechen funktionieren kann?

Es gibt doch unzählige Möglichkeiten, eine Regel zu interpretieren. Auch Wittgenstein war sich dieses Problems bewusst und beginnt dies folgendermaßen zu adressieren:

„Eine Reihe hat für uns ein Gesicht!“ – Wohl; aber welches? Nun doch das algebraische, und das eines Stücks der Entwicklung. Oder hat sie sonst noch eins? – „Aber in dem liegt doch schon alles!“ – Aber das ist keine Feststellung über das Reihenstück, oder über etwas, was wir darin erblicken; sondern der Ausdruck dafür, daß wir nur auf den Mund der Regel schauen und tun, und an keine weitere Anleitung appellieren. (PU 2009: §228)

Eine Sprecher*innengemeinschaft muss sich also auf einige geteilte Interpretationsmuster einigen. Dies reduzieren die unzähligen möglichen

Interpretationen von Regeln auf einige wenige, dadurch hält sich die Anzahl möglicher Interpretationen in überschaubarem Rahmen und macht Kommunikation somit überhaupt erst möglich. Wittgenstein behandelt diese Interpretationsmuster in Paragraph §240: „Es bricht kein Streit darüber aus (etwa zwischen Mathematikern), ob der Regel gemäß vorgegangen wurde, oder nicht. Es kommt darüber z. B. nicht zu Tätlichkeiten. Das gehört zu dem Gerüst, von welchem aus unsere Sprache wirkt (z. B. eine Beschreibung gibt).“ (PU 2009: §240) Kommunikation ist also ohne andauernden Streit über korrektes Regelfolgen möglich, weil es ein „Gerüst“ gibt, das Sprechen möglich macht. Hier tauchen starke Ähnlichkeiten mit dem Begriff der Lebenswelt bei Habermas auf, der die Lebenswelt wie folgt definiert: „Die Strukturen der Lebenswelt legen die Formen der Intersubjektivität möglicher Verständigung fest. Ihnen verdanken die Kommunikationsteilnehmer die extramundane Stellung gegenüber dem Innerweltlichen, über das sie sich verständigen können.“ (Habermas 2011: 192) Die Lebenswelt ist also ein Interpretationsmuster, das der Kommunikation vorausgeht und welche Kommunikation über intersubjektive Inhalte erst möglich macht, da es die Basis für Intersubjektivität darstellt.

Die Funktion dieses Gerüsts bei Wittgenstein scheint sich zumindest mit der Lebenswelt bei Habermas zu überschneiden. In zwei Paragraphen geht Wittgenstein darauf ein, wie die Struktur jenes Gerüsts zustande kommt und beschreibt dessen Funktion in der Sprache. Bezüglich des Zustandekommens schreibt Wittgenstein folgendes: „So sagst du also, daß die Übereinstimmung der Menschen entscheide, was richtig und was falsch ist? ‘ – Richtig und falsch ist, was Menschen sagen; und in der Sprache stimmen die Menschen überein. Dies ist keine Übereinstimmung der Meinungen, sondern der Lebensform.“ (PU 2009: §241) Die Übereinstimmung der Lebensformen ist also die Ursache für die Möglichkeit des Regelfolgens in der sprachlichen Praxis. Somit scheint die Funktion der Lebensform bei Wittgenstein dieselbe Funktion zu erfüllen, die der Begriff der Lebenswelt bei Habermas erfüllt. Gäbe es weder Lebensform noch Lebenswelt, könnten sich Sprecher*innen nicht über Inhalte austauschen, da die Grundlage zur Interpretation des Gesagten fehlen würde. Die Lebenswelt / Lebensform erzeugt die Möglichkeit der Kommunikation und Ausverhandlung von intersubjektiven Inhalten. Wittgenstein schreibt weiter über die Funktion der Lebensform:

Zur Verständigung durch die Sprache gehört nicht nur eine Übereinstimmung in den Definitionen, sondern (so seltsam dies klingen mag) eine Übereinstimmung in den Urteilen. Dies scheint die Logik aufzuheben; hebt sie aber nicht auf. – Eines ist, die Meßmethode zu beschreiben, ein Anderes, Messungsergebnisse zu finden und auszusprechen. (PU 2009: §242)

Habermas scheint diese Ansicht zu teilen und beschreibt die Lebenswelt im Kontrast zu den drei Welten folgendermaßen: Die Strukturen der Lebenswelt legen die Formen der Intersubjektivität möglicher Verständigung fest. (Habermas 2011: 192)

Diese Intersubjektivität wird gerade durch eine Übereinstimmung der Urteile erzeugt. Die Lebenswelt ist der Raum, in dem diese geteilten Urteilsmechanismen verortet sind, die wir durch eine gemeinsame Lebensform erhalten.

Habermas schreibt: „Die kommunikativ Handelnden bewegen sich stets innerhalb des Horizonts ihrer Lebenswelt; aus ihm können sie nicht heraustreten.“ (Habermas 2011: 191). Dem Gerüst bei Wittgenstein kommt eine ähnliche Rolle zu: Auch hier ist das Gerüst, die Übereinstimmungen der Lebensformen, die Voraussetzung für kommunikative Prozesse. Jedoch scheint Habermas die fundamentale Rolle der Lebenswelt als etwas, aus dem man nicht heraustreten kann, etwas, worüber man kaum sprechen kann, fundamentaler zu sehen als Wittgenstein, zumindest wenn man sich nur an die „Philosophischen Untersuchungen“ hält. Jedoch gibt es in „Über Gewissheit“ eine Stelle, in der Wittgenstein über das Irren nachdenkt, Gedanken enthält, die eine ähnliche Fundamentalität der Lebensform wie die der Lebenswelt andeuten. Er schreibt dazu:

Der Mensch kann sich unter gewissen Umständen nicht *irren*. (›Kann‹ ist hier logisch gebraucht, und der Satz sagt nicht, daß unter diesen Umständen der Mensch nichts Falsches sagen kann.) Wenn Moore das Gegenteil von jenen Sätzen aussagte, die er für gewiß erklärt, würden wir nicht nur nicht seiner Meinung sein, sondern ihn für geistesgestört halten. (ÜG 2020: 155).

Der Mensch kann sich also in gewissen Situationen nicht irren, denn ein Irrtum an gewissen Sätzen wäre kein Irrtum in der Art, wie wir das Wort gebrauchen, sondern eine Geistesstörung. Wittgenstein ergänzt dies noch mit folgender Aussage: „Damit der Mensch sich irre, muß er schon mit der Menschheit konform urteilen.“ (ÜG 2020: 156). Um sich also zu irren, muss eine Sprecher*in mit der Menschheit konform gehen. Hier schreibt Wittgenstein selbst nichts über die Lebensform, jedoch werde ich in diesem Aufsatz vertreten, dass dies nur bedeuten kann: Damit der Mensch sich irren kann, muss er mit der Lebensform, die den Begriff des Irrtums hervorgebracht hat, konform urteilen. Ein Irrtum ist also etwas, das in unserer Lebensform als möglicher Irrtum vorgesehen ist, also nicht über die Grenzen des Horizonts unserer Lebenswelt hinausgeht. Wer diesen Horizont überschreitet, begibt sich auf das Feld der Geistesstörung, er verliert somit alle kommunikativen Ansprüche und kann auch nicht mehr auf die drei Welten referenzieren; seine Aussagen sind kein Irrtum, sondern nicht Teil des Sprachspiels selbst. Diese Überlegung lässt sich an folgender Aussage Wittgensteins in *Über Gewißheit* plausibel begründen; er schreibt: „Vom Menschen, in Moores Sinne, zu sagen, er wisse etwas; was er sage, sei also unbedingt die Wahrheit, scheint mir falsch. – Es ist die Wahrheit nur insofern, als es eine unwankende Grundlage seiner Sprachspiele ist.“ (ÜG 2020: 403). Moore weiß diese Dinge, in denen man sich nicht irren kann, da man sonst „geistesgestört“ wäre, deshalb, weil sie Grundlage und Hintergrund des Sprachspiels sind. Etwas, das sich wohl nur durch eine fundamentale Stellung jener Erfahrungen in der Lebensform erklären lässt. Dieses Vorhandensein von fundamentalen Gegebenheiten, Erfahrungssätzen und Deutungsmustern teilt das Sprachspiel und, subsequent, die Lebensform mit der Lebenswelt von Habermas. Wie steht es aber um ihre Genese, finden sich auch im Zustandekommen der Lebenswelt und der Lebensform Ähnlichkeiten?

Habermas schreibt über die Zusammensetzung und den Prozess der Bildung der Lebenswelt folgendes:

Während sich a fronte dem Handelnden der situationsrelevante Ausschnitt der Lebenswelt als Problem aufdrängt, das er in eigener Regie lösen muß, wird er a tergo vom Hintergrund seiner Lebenswelt getragen, die keineswegs nur aus kulturellen Gewißheiten besteht. Dieser Hintergrund besteht auch aus individuellen Fertigkeiten, dem intuitiven

Wissen, wie man mit einer Situation fertig wird, und aus sozial eingelebten Praktiken, dem intuitiven Wissen, worauf man sich in einer Situation verlassen kann, nicht weniger als aus den trivialerweise gewußten Hintergrundüberzeugungen. (Habermas 2011: 204-205)

Die Lebenswelt besteht aus individuellen Fertigkeiten, Wissen, wie man mit Situationen umzugehen hat, sozialen Praktiken und auch den Überzeugungsätzen von Gruppenidentitäten. Gleichwohl können auch die Inhalte der drei Welten als Teil der Lebenswelt fungieren. Die Beziehung der Lebenswelt und der Inhalte des Sprechens ist reziproker Natur, denn sie befindet sich durch den Akt des Sprechens und die Praxen der sozialen Gruppe und Individuen im ständigen Wandel und wird durch den Kommunikationsprozess neu verhandelt, für den sie dann wieder den Hintergrund darstellt. Die Lebenswelt ist somit eine Sammlung von Überzeugungen, Wissen und geübten Praxen.

Um uns ein Verständnis des Begriffs der Lebensform bei Wittgenstein anzueignen, müssen wir etwas über die PU hinausgehen, da dort der Begriff nur viermal auftaucht und ungenügend erklärt wird. In Wittgensteins Nachlass lässt sich folgende Definition der Lebensform finden:

Statt des Unzerlegbaren, Spezifischen, Undefinierbaren, die Tatsache, daß wir so & so handeln, z.B., gewisse Handlungen strafen, den Tatbestand so & so feststellen, Befehle geben, berichten, z.B. Farben beschreiben, uns für die Gefühle der Andern interessieren. Das hinzunehmende, gegebene – könnte man sagen – seien Lebensformen. || seien Tatsachen des Lebens. (MS 133: 28r)

Lebensform bei Wittgenstein ist ein aus der Praxis folgendes Konstrukt, das jedem Menschen zugrunde liegt. Die Ähnlichkeit unserer Lebensform eröffnet uns die Ähnlichkeit unserer Interpretationsmuster, unserer Urteile und unseres Gerüsts. Wie bei Habermas ist auch die Lebensform bei Wittgenstein etwas, das aus der Praxis einer Gemeinschaft oder unseres eigenen Handelns erwächst. Wer aus den Urteilen der Lebenswelt hinaustritt, handle, als wäre er geistesgestört, wie Wittgenstein schreibt. Dennoch ist das, was wir tun, nicht festgeschrieben, sondern befindet sich im Wandel. Somit ist auch die Lebensform genau wie die Lebenswelt ein sich im Wandel befindlicher Begriff

und etwas, das sich auch durch Sprechakte zu verändern vermag. Daraus folgt, dass sowohl die Lebenswelt als auch die Lebensform die Grundlage für das Sprechen darstellt und dieses erst ermöglichen. Gleichzeitig ist aber sowohl die Lebenswelt als auch die Lebensform konstanter Veränderung unterzogen und auch durch den Akt des Sprechens selbst veränderbar.

3. Konklusion

Bei der Gegenüberstellung der Sprachphilosophie von Wittgenstein und Habermas fällt sofort der starke Fokus auf die sprachliche Praxis und den Akt des Sprechens als ein Instrument zur Kommunikation von Gruppen auf. Diese Ähnlichkeiten verdichten sich weiter, wenn man beginnt, die beiden Begriffe der Lebenswelt bei Habermas und den Begriff der Lebensform bei Wittgenstein zu vergleichen. Sie erfüllen beide die Rolle eines Hintergrunds oder Gerüsts, das das Sprechen erst ermöglicht. Auch in ihrer Struktur und Genese ähneln sie sich sehr stark, sind sie doch beide Produkt der Praxis und Überzeugungsätze einer Sprecher*innengemeinschaft. Ein Hintergrund, der ständig im Wandel ist und durch jeden Akt des Sprechens neuverhandelt, also entweder bestätigt oder verändert, wird.

Diese Ähnlichkeiten in Struktur und Funktion der beiden Begriffe sind es, die mich in dieser Arbeit zur Konklusion führen, dass es sich bei diesen Begriffen um verwandte und sich stark überlappende Begriffe handelt. Diese Überlappungen in der fundamentalen Rolle der Praxis des Lebens als auch in der Hervorbringung des Intersubjektiven öffnen Wittgenstein wohl auch für phänomenologische Analysen, da ja auch die hier behandelten Habermas'schen Konzepte Entlehnungen aus eben dieser sind. Eine Frage für zukünftige Untersuchungen kann nun sein, wie sich diese Phänomene des Intersubjektiven, des geteilten Erlebens, für Analysen unserer Sprache fruchtbar machen lassen. Ein mir mehr als interessant erscheinender Gedanke ist wohl die Möglichkeit, dadurch die Analyse der Sprache auf den Kopf zu stellen. Wittgenstein fragt immer nach dem Gebrauch eines Wortes im Sprachspiel, jedoch scheint es mir nun auch zielführend, die Lebenswelt, die Lebensform, die ein Sprachspiel hervorgebracht hat, zum Objekt der

Befragung zu machen. Dies könnte neue spannende Einsichten in den Gebrauch der Sprache liefern, Einsichten, die womöglich durch eine reine Analyse des Gebrauchs im Sprachspiel dem Blick verwehrt bleiben würden.

References

Beyer, Christian (2020) "Edmund Husserl", in: *Stanford encyclopedia of philosophy*. URL: <https://plato.stanford.edu/entries/husserl/#EmpIntLif>

Habermas, Jürgen (2011) *Theorie des kommunikativen Handelns. 2, Zur Kritik der funktionalistischen Vernunft*, 8. Auflage. Frankfurt am Main: Suhrkamp.

Wittgenstein Archives at the University of Bergen (Hg.) 2000: *Wittgenstein's Nachlass. Text and Facsimile Version*, The Bergen Electronic Edition, Oxford: Oxford University Press. Als MS, TS oder D mit der jeweiligen Manuskript-, Typoskript- oder Diktatnummer und der Seitenzahl zitiert.

Wittgenstein, Ludwig (2009) *Philosophical Investigations / Philosophische Untersuchungen* Hg. von P. M. S. Hacker und Joachim Schulte, übersetzt von G. E. M. Anscombe, P. M. S. Hacker und Joachim Schulte. New York: Wiley-Blackwell. Deutsch mit englischer Übersetzung, zit. mit Sigle: PU 2009.

Wittgenstein, Ludwig (2020) *Über Gewißheit*, Hg. von G. E. M. Anscombe und G. H. von Wright. Neu durchgesehen mit Rush Rhees. 15. Auflage, Frankfurt am Main: Suhrkamp. Mit Sigle: ÜG 2020.

Wittgenstein über Gesichtsraum und Grammatik

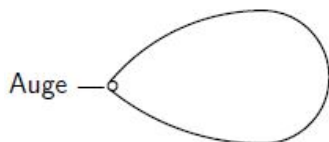
Maximilian Philipps (Mannheim, Germany)

Abstract

Ein Thema, dem sich Wittgenstein über nahezu die gesamte Phase seines Schaffens hinweg widmet, ist das der Beziehung zwischen Subjekt und Gesichtsraum. Erstmals kommt er hierauf in einer zentralen Passage des *Tractatus* zu sprechen, um dann das Thema in seiner mittleren und späten Schaffensphase wieder und wieder aufzugreifen. Eine beträchtliche Anzahl der entsprechenden Textstellen ist nicht in gedruckter Form, sondern lediglich über die *Bergen Electronic Edition* verfügbar, was die bisher eher spärliche Rezeption erklären könnte. In diesem Aufsatz will ich Wittgensteins Überlegungen rekonstruieren und dabei zeigen, dass seine Diskussion des Begriffs des Gesichtsraums nicht primär auf Phänomene der visuellen Wahrnehmung abzielt, sondern auf die Grammatik dieses Begriffs.

Im *Tractatus* thematisiert Wittgenstein den Begriff des Gesichtsraums, bzw. den des Gesichtsfelds (beide Ausdrücke verwendet er synonym) in Abschnitt 5.6331:

Das Gesichtsfeld hat nämlich nicht etwa eine solche Form:



(TLP 5.6331)

Worauf Wittgenstein in seiner Diskussion dieser Zeichnung zunächst aufmerksam machen will, ist ein irreführender Aspekt in ihr: Das Auge ist zwar als Sinnesorgan dafür notwendig, dass das Subjekt überhaupt ein Gesichtsfeld haben kann, aber es ist selbst kein gesehener Bestandteil davon – das Auge, bzw. das Subjekt, kann sich (außer im Spiegel) nicht selbst betrachten: „Und nichts am Gesichtsfeld läßt darauf schließen, daß es von einem Auge gesehen wird.“ (TLP 5.633) Die Diskussion des Gesichtsfelds im *Tractatus* ist jedoch kein thematischer Selbstzweck. Vielmehr soll sie illustrieren, wie es sich mit der Beziehung des Subjekts zur Welt (nicht) verhält. (Vgl. TLP 5.633) In der Konzeption des *Tractatus* ist das Ich, das „denkende, vorstellende Subjekt“ (TLP 5.631) explizit *kein* „Teil der Welt“, sondern die „Grenze“ der Welt:

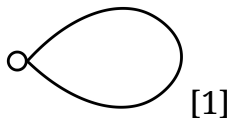
Das philosophische Ich ist nicht der Mensch, nicht der menschliche Körper, oder die menschliche Seele, von der die Psychologie handelt, sondern das metaphysische Subjekt, die Grenze – nicht ein Teil – der Welt. (TLP 5.641)

Die Zeichnung in TLP 5.6331 ist keine adäquate Darstellung der Relation von metaphysischem Subjekt zur Welt. Erstens, weil sie durch die Inklusion des Auges, welches metaphorisch für das metaphysische Subjekt steht, suggeriert, dieses Subjekt *sei* Teil der Welt. Und zweitens, weil sie suggeriert, das Feld habe eine Begrenzung. Freilich hat die visuelle Wahrnehmung im physikalischen Sinne ihre Grenzen, da man z.B. Gegenstände nur bis zu einer bestimmten Entfernung erkennen kann, aber insofern die Darstellung sich nicht auf den „menschlichen Körper“ (und dessen Wahrnehmungsapparat) bezieht, sondern auf das *metaphysische Subjekt* und dessen Verhältnis zur Welt bezieht, verhält es sich anders: „Der Tod ist kein Ereignis des Lebens. Den Tod erlebt man nicht. [...] *Unser Leben ist ebenso endlos, wie unser Gesichtsfeld grenzenlos ist.*“ (TLP 6.4311, Hervorhebung M.P.) Das Gesichtsfeld, metaphorisch verstanden als perspektivische Beziehung des metaphysischen Subjekts zur Welt, hat deshalb keine Grenze, weil das Subjekt nicht aus seiner Perspektive heraustreten kann; es kann sich nicht selbst zum Gegenstand der eigenen Wahrnehmung machen. Die Darstellung und Behauptung einer Grenze ist nur da sinnvoll, wo eine Verortung auf beiden Seiten dieser Grenze möglich ist: Ebenso wenig, wie das metaphysische Subjekt eine Wahrnehmung jenseits der eigenen Perspektive erleben kann, kann es Erfahrungen jenseits seines eigenen Lebens machen.

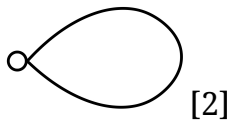
Eine adäquate Darstellung der Perspektive des metaphysischen Subjekts (welches man auch als *transzendentes* Subjekt verstehen kann, da es die Bedingung der Möglichkeit der Wahrnehmung ist) und seiner Beziehung zur Welt dürfte also weder den Punkt enthalten, der das Subjekt, bzw. Auge, symbolisiert, noch eine Grenzlinie des Gesichtsfeldes. Es bliebe also nichts mehr übrig. Folgerichtig kommt Wittgenstein 1930 zum radikalen Schluss: „Nein, ein sichtbares Bild des Gesichtsbildes kann man nicht machen.“ (MS 108: 39f. Er bezieht sich dabei auf eine Darstellung Ernst Machs, auf die ich noch weiter unten zu sprechen komme.) Im Nachfolgenden will ich Wittgensteins Überlegungen, die zu diesem Kulminationspunkt hinführen, rekonstruieren.

Die erste Anmerkung zum Gesichtsfeld, die Wittgenstein nach der Abschrift der Urfassung des *Tractatus* im Jahr 1918 verfasste, entstammt einem Brief, den er am 23.04.1922 an C.K. Ogden, den Übersetzer des *Tractatus*, schrieb. In diesem geht es um Korrekturwünsche die Urfassung betreffend, und unter anderem nimmt Wittgenstein hier explizit Bezug auf die Zeichnung des Gesichtsfelds in TLP 5.6331:

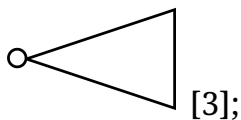
As to



you will see from my correction as well as from the German that the prop[osition] had been mutilated. The figure s<c>hould be like this



and not



because this [2] is how people often imagine the shape of the field of vision to be.

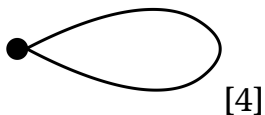
(CCO 1973, S. 20, meine Nummerierung, M.P.)

Zwei Aspekte sind hier von Interesse. Der erste, offenkundige, besteht darin, dass Wittgenstein mit der dreieckigen Zeichnung in der Urfassung nicht zufrieden war und auf die ovale Form bestand. (Ob es tatsächlich der Fall ist, dass Menschen sich das Gesichtsfeld typischerweise wie [2] vorstellen, wie von Wittgenstein angenommen, sei dahingestellt.) Der zweite, weniger offenkundige Aspekt besteht darin, dass die bis heute in den Druckfassungen des *Tractatus* verwendete Grafik in TLP 5.6331 in einem kleinen Detail von Wittgensteins gewünschter Darstellung abweicht: Der Punkt, der das Subjekt (bzw. dessen Auge) symbolisiert, dessen Gesichtsfeld die Grafik darstellen soll, liegt in Wittgensteins Variante der ovalen Zeichnung *außerhalb* des Ovals,

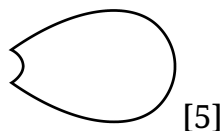
welches das Gesichtsfeld symbolisiert, nicht innerhalb. (Die Folgen dieses Umstands für die *Tractatus*-Exegese diskutiert Bazzocchi 2014 ausgiebig.) Als Wittgenstein darstellen wollte, welche Form das Gesichtsfeld nicht hat – eben eine, in der eine Grenze erkennbar ist, – hatte er also von Beginn an eine Darstellung im Sinn, in der der symbolische Punkt außerhalb des Ovals liegt. Dass er bis heute in den Druckversionen des *Tractatus innerhalb* des Ovals liegt, passt nicht zu Wittgensteins Aussage in (u.a.) TLP 5.631, dass das transzendente Subjekt kein Teil der (von ihm wahrgenommenen) Welt ist, sondern im metaphysischen Sinne außerhalb der Welt steht.

In TS 212 kommt Wittgenstein explizit auf seine ursprüngliche Zeichnung aus TLP 5.6331 zu sprechen:

Es ist seltsam, daß ich geschrieben habe, der Gesichtsraum hat nicht die Form



und nicht er habe nicht die Form



und daß ich das erste geschrieben habe, ist sehr bezeichnend.

(TS 212: 436, inhaltlich gleich auch in MS 112 enthalten.)

Wie ich zeigen will, stellt die untere, revidierte Darstellung [5] des Gesichtsfelds sozusagen die grafische Pointe hinter dem in TLP 5.633 angestoßenen und mit TLP 5.6331 illustrierten Gedanken über das Verhältnis zwischen Subjekt und Welt dar. Symbolisierte Wittgenstein im Brief an Ogden das Subjekt noch mit einem kleinen Punkt, von dem das Gesichtsfeld ausgeht, betont die revidierte Zeichnung [5] mit der kleinen Aussparung am linken Ende nun die *Abwesenheit* eines Subjektes. Wie oben diskutiert will Wittgenstein mit seinen Bemerkungen über die grafische Darstellung des Gesichtsfelds im *Tractatus* darauf hinaus, dass das Gesichtsfeld *richtig*

verstanden keine Grenze hat – aber es ist nicht dieser Aspekt, um den es ihm hier in TS 212 geht. Die Zeichnung in TLP 5.6331, bzw. die richtiggestellten Version [1]/[2] oder [4], ist eine negative Charakterisierung des Gesichtsfelds; eine zeichnerische Art der Beschreibung, wie es *nicht* *verstanden* werden sollte: Als etwas, das eine Grenze hat. Indem sie die Abwesenheit des Subjektes betont, zeigt Zeichnung [5] mit der Aussparung jedoch einen Aspekt auf, der gemäß Wittgensteins Verständnis des Gesichtsfelds dem Gesichtsfeld *essentiell* ist: Die Abwesenheit des Subjekts als Abwesenheit eines „Besitzers“ des Gesichtsfelds. Wittgenstein macht dies in den zur gleichen Zeit entstandenen *Philosophischen Bemerkungen* explizit: „Der Gesichtsraum hat wesentlich keinen Besitzer. [...] Das Wesentliche ist, daß die Darstellung des Gesichtsraums ein Objekt darstellt und keine Andeutung eines Subjekts enthält.“ (PB 71) Mit anderen Worten: Eine angemessene Darstellung des Gesichtsraums würde lediglich *Gesehenes* abbilden, aber nicht das sehende Subjekt, weder als symbolischen Punkt, noch durch eine angedeutete Sichtfeldgrenze wie in Ernst Machs „Selbstportrait“, s.u.. Eine solche Darstellung wäre nutzlos, da sie für den uninformierten Betrachter nicht mehr *als Darstellung des Gesichtsraums* erkennbar wäre. Wittgenstein merkt oben an, es sei „seltsam“ und „sehr bezeichnend“, dass er sich zunächst für eine Zeichnung wie [4] entschieden habe. Meines Erachtens ist dies eine Reflexion auf den Umstand, dass man intuitiv zu Darstellungen wie [4] geneigt ist, welche das Subjekt inkludieren, und eben nicht zu einer wie [5], was Hand in Hand mit dem irreführenden Bild von Gesichtsraum und dessen „Besitzer“ geht.

Wenige Jahre später kommt er erneut auf TLP 5.6331 zurück, sich selbst grob zitierend:

„Nichts im Gesichtsfeld deutet darauf hin etc.“ (Log. Phil. Abh.) Das heißt sozusagen: Du wirst vergebens nach dem Sehen [sic!, vermutlich „Sehenden“, M.P.] ausschauen. Er ist nirgends im Gesichtsraum zu finden. Aber die Wahrheit ist: Du tust nur, als suchtest du nach einem Etwas, nach einer Person im Gesichtsraum, die nicht da ist. (MS 116: 235)

Das Hinzuziehen weiterer Passagen macht klar, wieso Wittgenstein das Bild von Gesichtsraum und zugehörigem Besitzer zurückweist und auf welcher Ebene er das zugrunde liegende philosophische Problem verortet sieht: Auf

Ebene der Sprache und des Sprachgebrauchs, und nicht etwa auf der physikalischen Ebene der visuellen Wahrnehmung. Es ist nicht sein Ansinnen, eine möglichst adäquate grafische Darstellung des Gesichtsraums zu liefern – wie bereits dargelegt, hält er dies für unmöglich. Worum es ihm in seiner Diskussion des Gesichtsraums in seiner mittleren Periode wirklich geht, ist das Potenzial der Sprache, den Sprachverwender in die Irre zu führen. Unzweifelhaft klar wird dieser Punkt in den *Philosophischen Bemerkungen*:

Schon das Wort ›Gesichtsraum‹ ist für unseren Zweck ungeeignet, denn es enthält eine Anspielung auf ein Sinnesorgan, die für den Raum ebensowenig wesentlich ist, als es für ein Buch wesentlich ist, daß es einem bestimmten Menschen gehört; und es könnte sehr irreführend sein, wenn es in unserer Sprache so eingerichtet wäre, daß wir in ihr kein Buch bezeichnen könnten, außer durch seine Beziehung zu einem Besitzer. (PB 73)

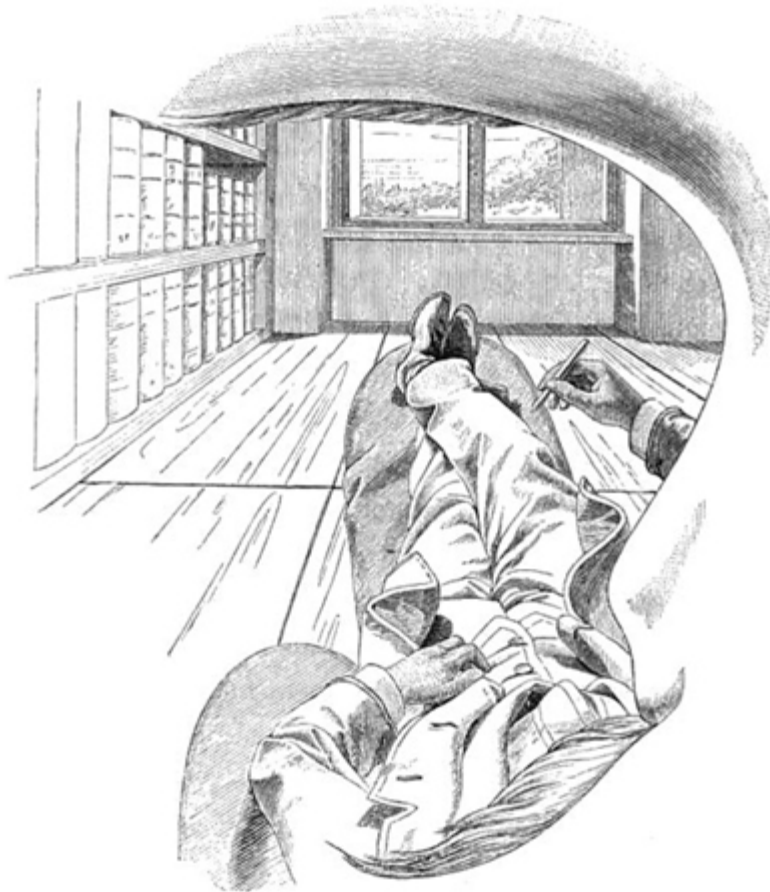
Der fragliche Zweck, den Wittgenstein verfolgt, ist das Ausräumen philosophischer Irrtümer, die durch Missdeutung der Sprache entstehen. (Wie der, visuelle Eindrücke seien etwas, über das man in einer Terminologie von Gesichtsräumen und deren jeweiligen Besitzern spräche, s.u.) So erklärt er in MS 107: 213: „Die philosophische Aufgabe mit Beziehung auf den Gesichtsraum besteht eben – wie immer – nur darin falsche philosophische Theorien über ihn zurückzuweisen.“ Und weiterhin:

Niemand kann uns unseren Gesichtsraum näher kennen lehren. Aber wir können seine sprachliche Darstellung übersehen lernen. [...] Wenn ich sage: Hier sind wir an der Grenze der Sprache, so scheint || klingt das immer, als wäre hier eine Resignation nötig, während im Gegenteil volle Befriedigung eintritt, da keine Frage übrigbleibt. [...] Alles, was von Interesse || von logischem Interesse, was wir liefern können, ist die Beschreibung der Sprache. (MS 110: 98f)

Dies wirft die Frage auf, worin konkret der Irrtum besteht, wenn man den Gesichtsraum als etwas versteht, das im „Besitz“ des wahrnehmenden Subjektes ist. Welche Einblicke gewinnen wir, wenn wir hier die „sprachliche Darstellung übersehen lernen“ und wo liegt die „Grenze der Sprache“? Verfolgen wir Wittgensteins Gedanken weiter:

Wenn wir vom Gesichtsraum reden, so werden wir leicht zu der Vorstellung verführt, als wäre er eine Art von Guckkasten, der jeder mit || vor sich herumtrüge. D.h. wir verwenden dann das Wort ‚Raum‘ ähnlich, wie wenn wir ein Zimmer einen Raum nennen. In Wirklichkeit aber bezieht sich doch das Wort ‚Gesichtsraum‘ nur auf eine Geometrie, ich meine, *auf einen Abschnitt der Grammatik unserer Sprache*. In diesem Sinne gibt es keine ‚Gesichtsräume‘ die etwa jeder seinen Besitzer hätten. (Hervorhebung M.P.) (TS 213: 463)

Die „Verführung“, vor der Wittgenstein warnt, besteht also darin, ausgehend vom Begriff „Raum“ den Gesichtsraum als einen physikalischen, visuell wahrnehmbaren Raum zu verstehen, der einem Subjekt im possessiven Sinne zukommt. Richtig verstanden, so seine Auffassung, ist eine Diskussion des Gesichtsraums aber keine Diskussion eines physikalischen Raums, sondern die Diskussion eines bestimmten Bereich der Sprache, der seine eigenen, charakteristischen Strukturen und Regeln („Grammatik“) besitzt. An anderer Stelle führt er weiter über den Begriff der Geometrie aus: „Die Geometrie des Gesichtsraums ist die Syntax der Sätze, die von den Gegenständen im Gesichtsraum handeln“ (MS 107: 213), und: „[D]ie Geometrie ist hier einfach Grammatik.“ (MS 108: 95) Wittgenstein versteht ein sinnvolles Reden über den Gesichtsraum also eigentlich als Reden über einen „Sprachraum“ (meine Formulierung). Als solcher kann er keinen Besitzer haben, und als solcher kann er auch nicht auf empirischem Wege erschlossen werden: „Die Untersuchung [des Gesichtsraums, M.P.] ist keine im Sinn einer physikalischen oder psychologischen Untersuchung.“ (BT 444) Vielmehr muss die Untersuchung auf Ebene der Sprachverwendung stattfinden; die „Geometrie [des Gesichtsraums, M.P.] ist Grammatik und die Untersuchung eine grammatische Untersuchung.“ (Ebd.) Es passt ins Bild, dass er in einer Anmerkung zu Ernst Machs Darstellung des Gesichtsraums (s.o., MS 108: 39f) feststellt, ein „sichtbares“ (also gezeichnetes) Bild des Gesichtsraums könne grundsätzlich nicht erstellt werden, und in Machs Darstellungsversuch eine kategoriale „Verwechslung“ sieht: Korrekt, also nicht-physikalisch verstanden, ist der Gesichtsraum nicht die Art von Ding, die sich sinnvoll grafisch darstellen lässt.



Figur 1.

(Mach 1886: 14)

In seiner frühen Periode liefen Wittgensteins Überlegungen bezüglich der Darstellbarkeit darauf hinaus, dass sich der Gesichtsraum, metaphorisch verstanden als perspektivischer Zugang des metaphysischen (und transzendentalen) Subjekts zur Welt, deshalb nicht darstellen lässt, weil das Subjekt selbst kein Teil der Darstellung sein darf (da kein Teil der Welt) und der Gesichtsraum keine Grenze hat, da die erstepersonale Perspektive des metaphysischen Subjekts kein „Jenseits“ kennt. In der mittleren und späten Periode verhält es sich nun anders: Ein korrektes Verständnis des Gesichtsraums bedeutet eigentlich, die Sprachgrammatik des Redens über visuelle Wahrnehmung zu verstehen. Wo es hier eine Grenze gibt, so ist es keine, die sich grafisch darstellen ließe – sie ist im Sprachgebrauch selbst zu suchen, wie beide folgenden Passagen verdeutlichen:

Es ist unsinnig zu sagen ‚ich sehe die Dinge || diesen Gegenstand im Gesichtsraum‘. Im Gegensatz wozu? Ist es denkbar, daß ich sie höre, oder

daß ein Anderer sie || ihn höre, oder daß ein Anderer ihn sieht? (TS 212: 412)

„Aber kann nicht ich in meinem Gesichtsraum eine Landschaft, und Du in dem Deinen ein Zimmer sehen?“ – Nein, – „ich sehe in meinem Gesichtsraum“ ist Unsinn. Es muß heißen „ich sehe eine Landschaft und Du etc.“ – und das wird nicht bestritten. Was uns hier irreführt, ist eben das Gleichnis vom Guckkasten, oder etwa von einer kreisrunden weißen Scheibe, die wir gleichsam als Projektionsleinwand mit uns trügen, und die der Raum ist, in dem das jeweilige Gesichtsbild erscheint. (TS 212: 706)

Eine grammatische Betrachtung offenbart, dass solche Redeweisen jenseits der Grenze des sinnvollen Sprachgebrauchs liegen, also „Unsinn“ sind. Eine Aussage wie „Ich sehe etwas *im Gesichtsraum*“ suggeriert „Gegensätze“, die grammatisch prima facie ausgeschlossen sind, nämlich ein Sehen *außerhalb* des Gesichtsraums, das haben visueller Eindrücke *einer anderen Person* in „meinem“ Gesichtsraum, sowie das Haben von nichtvisuellen Sinneseindrücken im Gesichtsraum. Wieder betont Wittgenstein, dass das Befürworten einer solchen Redeweise das Resultat einer Irreführung ist: Aus der falschen Auffassung des Gesichtsraums als physikalischem Raum erwächst die Vorstellung des Subjekts, das „seine“ visuellen Eindrücke in „seinem“ Gesichtsraum hat, und, wie ich nun abschließend zeigen will, letztlich auch die für die *Philosophischen Untersuchungen* so zentrale Idee der vermeintlichen Privatheit von Empfindungen.

In PU 398 führt Wittgenstein einen Dialog mit einem fiktiven Diskussionspartner, welcher vorbringt: „Aber wenn ich mir etwas vorstelle, oder auch wirklich Gegenstände *sähe*, so *habe* ich doch etwas, was mein Nachbar nicht hat.“ Worauf Wittgenstein zunächst antwortet: „Ich verstehe dich. Du willst um dich schauen und sagen ›Nur *ich* habe doch Dieses.‹“ Damit steht Wittgensteins Opponent stellvertretend für die cartesianische Auffassung, nach welcher die visuellen Eindrücke des Subjekts in „seinem“ Gesichtsraum sozusagen in seinem privaten, „mentalenen Besitz“ sind. Es folgt die Zurückweisung dieser Position:

Du redest (wenn du z.B. im Zimmer sitzt) vom ›visuellen Zimmer‹. Das, was keinen Besitzer hat, ist das ›visuelle Zimmer‹. Ich kann es so wenig

besitzen, als ich darin umhergehen, oder es anschauen, oder darauf zeigen kann. Es gehört insofern nicht mir an, als es niemand anderm angehören kann. (Ebd.)

Wittgensteins negative Charakterisierung des „visuellen Zimmers“ ist identisch mit der des Gesichtsraums: Es hat keinen Besitzer und darf nicht analog zu einem physikalischen Raum verstanden werden. (Man beachte die Ähnlichkeit dieser Passage zu TS 213: 463, s.o.) Der Opponent begeht einen Kategorienfehler, wenn er dem „visuellen Zimmer“, bzw. dem Gesichtsraum, zuspricht, es könne im gleichen Sinne etwas enthalten, wie das reale Zimmer: So wie er der Besitzer des Sessels in seinem physischen Zimmer sein könne, so sei er der Besitzer der *visuellen Eindrücke* des Sessels, nur eben im mentalen, visuellen Zimmer. Dieser Irrtum ist das Resultat des Vermischens der Grammatik der Rede über das Haben von Sinnesempfindungen mit der über das Besitzen von Gegenständen.

References

Bazzocchi, Luciano (2014) „A Significant ‘False Perception’ of Wittgenstein’s Draft of the Mind’s Eye”, *Acta Analytica* 29(2), 255 – 256.

Mach, Ernst (1886) *Die Analyse der Empfindungen*, Jena: Verlag von Gustav Fischer.

Wittgenstein, Ludwig (1964) *Philosophische Bemerkungen*, Frankfurt am Main: Suhrkamp.

Wittgenstein, Ludwig (1973) *Letters to C.K. Ogden*. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1984) *Werksausgabe Band 1. Tractatus logico-philosophicus / Tagebücher 1914-1916 / Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp.

Wittgenstein, Ludwig (1984a) *Werksausgabe Band 5. Das Blaue Buch / Eine philosophische Betrachtung (Das Braune Buch)*, Frankfurt am Main: Suhrkamp.

Wittgenstein, Ludwig (BEE, 2000) *Wittgenstein's Nachlass: The Bergen Electronic Edition*, Oxford University Press: Oxford.

Wittgenstein, Ludwig (2020) *The Big Typescript*, 2. revidierte Auflage, Frankfurt am Main: Vittorio Klostermann.

Acknowledgement

Ich bedanke mich bei Wolfgang Freitag, Helge Rückert, und Nadja-Mira Yolcu für ihre hilfreichen Kommentare.

Non-naturalist Moral Causation

Audrey Powers (New Brunswick, USA)

Abstract

If we are moral realists interested in moral causation, it is assumed that we had better go in for naturalism over non-naturalism, as naturalism lets us account for moral causation while non-naturalism doesn't. I argue that this assumption is false. According to widely-accepted difference-making frameworks for causation and modifications to these frameworks meant to capture causal claims correctly, moral facts may be causally efficacious in plausible non-naturalist ontologies. So non-naturalists are in a better position than we might expect when it comes to positing moral causation. The upshot is that, if we are choosing what view to take of the metaphysics of morality, moral causation is not a good reason to choose naturalism over non-naturalism, or vice versa.

1.

It seems that moral facts can cause natural facts. When my friend breaks her promise to pick me up at the airport – when she acts morally wrongly in this manner – I feel that something bad went on as I wait for a taxi. The moral fact that it was wrong of her to keep her promise caused the natural fact that I felt that something bad went on.

Of course, not everybody thinks this is what's happening (e.g., Audi 1993, 1997, Dworkin 1996, Thomson 1996, McGinn 1997, Slutsky 2001). Maybe my feelings are caused not by the moral itself but instead, for example, my beliefs about the moral (Harman 1977). But let's say we want to hold that the moral really is causally efficacious. How shall we do this?

Well, one way we could do this is by picking a metaethical view that accommodates moral causation. And, we might think, non-naturalist moral realism is not such a view. The assumption is that, if we are moral realists interested in moral causation, we had better go in for naturalism (see Majors 2003, Cuneo 2006, McBrayer 2010, Stringer 2021), on which moral facts are causally efficacious like other natural facts. We had better not go in for non-naturalism, on which moral facts, being in some important sense different, are presumably not causally efficacious like natural facts (Cuneo 2006, Enoch 2017, Stringer 2021).

I argue that the above assumption is false. According to widely-accepted difference-making frameworks for causation and modifications to these

frameworks meant to capture causal claims correctly (see Lewis 1973, Woodward 2003, 2015, Yablo 1992, Loewer 2007a, Campbell 2010, Papineau 2013, Albert 2015, Ismael 2016, Blanchard 2020, Woodward 2020, Zhong 2020a,b, Vaassen forthcoming), moral facts may cause natural facts on non-naturalist views.

The paper proceeds as follows. First (§2), I briefly treat the distinction between naturalism and non-naturalism, and explain why it seems that there will not be moral causation on non-naturalism. Next (§3,4), I argue that, *contra* the consensus in the literature, non-naturalists may have moral causation after all. I conclude (§5) by discussing the upshots of these claims in the debate between naturalists and non-naturalists.

2.

Very roughly, non-naturalist moral realists (hereafter ‘non-naturalists’) think that, in some meaningful sense, moral facts exist and are “not reducible to or entirely grounded in other, non-moral facts, properties, and objects,” i.e., natural facts, properties, and objects (Enoch, 2017:30) – moral facts are, instead, ontologically special, *sui generis*, or different.

We may contrast these views with those of naturalists, who think that moral facts exist and are ontologically part of ‘the natural,’ i.e., that which scientific claims quantify over (Darwall et al. 1992, Smith 1994, Shafer-Landau 2003, Copp 2007).

It seems like naturalism can accommodate moral causation while non-naturalism cannot. Take three independently plausible principles:

1. **Causal Closure:** Every natural fact has a sufficient natural cause.
2. **Exclusion:** If an effect has a sufficient natural cause, no non-natural cause can also be a cause of the same effect, unless that effect is overdetermined.
3. **Non-Overdetermination:** There is no systematic overdetermination in our picture of moral causation.

If we combine these principles, it appears there is no room for moral causation on non-naturalism. Moral and natural causes compete for their effects, and moral causes seem to lose. Take my thinking something bad went on when my friend breaks her promise. My thinking this is a natural fact. So it must have a sufficient natural cause. This sufficient natural cause cannot be the moral fact that it is wrong that she broke her promise, because on non-naturalism, moral facts aren't natural facts. And this moral fact cannot be a cause at all, because it would overdetermine its effect if it were causally efficacious alongside the natural cause.

This is an exclusion argument, of the kind we are familiar with from the mental causation literature (e.g., Malcolm 1968, Kim 1998, 1989, 2005, Papineau 2002, Bennett 2003). For explicit comparisons between such exclusion arguments and arguments against moral causation, we can look to Majors (2003:133), Oddie (2005:187-91), Cuneo (2006:157), and Viggiano (2020:2942-3).

It is not just the enemies of non-naturalism who endorse this exclusion argument. Non-naturalists do as well (e.g., Thomson 1996). Enoch notes that often, “non-naturalists think of moral facts as causally inert (for causal powers are arguably among the signs of the natural)” (2017:30).

Of course, I think this argument is mistaken. We may have non-naturalism and moral causation too.

3.

Here, I briefly explicate a difference-making framework for causation, and explain how this framework delivers moral causation on non-naturalism.

The thought driving difference-making accounts is quite intuitive. Here's a rough formulation:

Difference-Making: X is a cause of Y if and only if a change in X, holding all else fixed, would result in a change in Y.

I'm borrowing some language from interventionists (see Woodward 2003, 2008, 2015, Campbell 2010, Raatikainen 2010, Zhong 2020a,b), but we find this

general idea in Lewisian counterfactual (Lewis 1973), thermodynamical (Loewer 2007a, Albert 2015, Ismael 2016), and probabilistic (Papineau 2022) accounts of causation.

Of course **Difference-Making** is a very rough sketch – I am ignoring a great deal of complexity in how these accounts are actually formulated. Depending on the account we pick, we will need to say something extra about, e.g., time asymmetry or similarity rankings or interventions or other moving parts. We will certainly have to say more about what it means to hold all else fixed (e.g., Goodman 1973, Lewis 1979, Bennett 2003), so, in particular, we should take this to be a useful shorthand.

But, for my purposes, we can fill in most of these details as we like. In this manner I remain neutral as to the correct account of causation, and avoid depending on particular features of any account – my goal is to argue for non-naturalist moral causation generally, not non-naturalist moral causation on any particular account. Also, as a rough sketch, **Difference-Making** clearly needs modification – and, I argue, such modification is what gives us non-naturalist moral causation.

Here's how. Depending on what we hold fixed when testing for causation, **Difference-Making** has trouble capturing certain cases of higher-level causation (Shapiro and Sober 2007, Baumgartner 2010, Vaassen forthcoming). There exist some higher-level facts that are metaphysically necessitated by, but not identical to, lower-level facts. My doing a cartwheel is metaphysically necessitated by a set of facts about the positions of particles over time, but my doing a cartwheel isn't identical to those facts. Say my doing a cartwheel causes a vase to break when I hit it with my foot mid-rotation. We'd want to test for causation by asking whether a change to my cartwheeling, holding all else fixed, would have resulted in a change in the vase breaking. But we will come to a problem if some of the 'all else' we'd have to hold fixed is the facts that metaphysically necessitate my cartwheeling. It isn't metaphysically possible to change my cartwheeling without changing these facts. So, if we're taking these facts as part of the 'all else' to be held fixed, there aren't any possible changes to my cartwheeling holding all else fixed, and therefore my cartwheeling can't have caused the vase to break. But this is obviously wrong – my cartwheeling did cause the vase to break.

To ensure that **Difference-Making** gives us good results, we will just have to exclude such lower-level facts from the category of things to hold fixed when testing for causation. According to difference-makers, this is a principled move. The point of holding things fixed is to isolate potential causes from their epiphenomena to determine causation. For example: smoking causes tar-stained fingers and cancer. Actual changes to tar-stained fingers might correlate with differences in cancer risk, but only because such changes will likely – but very contingently – also be changes to smoking. However, if we change tar-stained fingers while holding smoking fixed, we will see no change in cancer risk. Here is the difference between this case and the case of my cartwheeling. It is metaphysically impossible to change any cause without also changing the things that metaphysically necessitate this cause (or the things the cause metaphysically necessitates). So actual changes to my cartwheeling will necessarily be changes to the things that metaphysically necessitate it. Even if we do not hold fixed the things that metaphysically necessitate my cartwheeling when changing my cartwheeling, we run no risk of picking up on pesky contingent correlations as in the smoking case. So: we may exempt things that metaphysically necessitate, or are metaphysically necessitated by, potential causes from being held fixed. Woodward (2014, 2015) and Vaassen (forthcoming) endorse such a move, making things look as follows:

Modified Difference-Making: X is a cause of Y if and only if a change in X, holding all else fixed except that which stands in a metaphysical necessitation relation to X, would result in a change in Y.

Modified Difference-Making should now deliver better causal claims in cases not having to do with moral causation.

Here, then, is how non-naturalists looking for moral causation make their move.

Either non-natural moral facts are fully grounded by a set of natural facts, or they aren't. Together, these options exhaust logical space.

Most non-naturalists think the latter – recall Enoch's characterization of the non-naturalist's moral facts as “not entirely grounded in other, non-moral facts, properties, and objects” (2017 p.30). But let's first look at those who think the former (Leary 2017, Berker 2018, perhaps Shafer-Landau 2003, and

discussion thereof in McPherson and Plunkett 2023). If these accounts are right, then the non-naturalist's work when it comes to moral causation is pretty much done. Take the common assumption that grounds metaphysically necessitate what they ground (*pace* Skiles 2015). If so, moral facts are metaphysically necessitated by natural facts.

With these considerations in mind, we can straightforwardly plug cases into **Modified Difference-Making** to see how non-naturalist moral causation might go. Consider my friend breaking her promise to pick me up at the airport. The moral fact that my friend acted wrongly causes my thinking that something bad went on if and only if a change to this moral fact, holding all else fixed except that which metaphysically necessitates or is metaphysically necessitated by the moral fact, would have resulted in a change in my thinking that something bad went on. When testing for causation, we should not consider the metaphysically impossible world where the metaphysically necessitating natural facts (that my friend drove her past the airport without stopping, that she had a certain mental state, that certain particles were arranged in a certain way, etc.) are the same but there is no moral fact, to see if I still would have thought that something bad went on. We should instead consider the metaphysically possible world where neither the metaphysically necessitating natural facts nor the moral facts obtain. And in such a world, presumably, I would not have thought that something bad went on. So the moral fact that it was wrong for my friend to break her promise caused my thinking that something bad went on. *Voilà* – moral causation on non-naturalism.

What about non-naturalists who don't think non-natural moral facts are fully grounded by natural facts? This is a more common position, endorsed by Enoch (2011, 2019), Scanlon (2014), Rosen (2017a,b,c), and Fogal and Risberg (2020), who think that moral facts are partially grounded by natural facts, but moral laws do some further work to make moral facts obtain.

Now, either these laws hold by metaphysical necessity, or they do not. The standard view is that they do – we generally think moral laws are metaphysically necessary (Bhagal, forthcoming).

It turns out it's easy to posit moral causation on the standard view. (We may also posit moral causation on the non-standard view, if we make another well-motivated adjustment to **Modified Difference-Making**, but I don't have space to make this argument here.) If moral laws are metaphysically necessary, they, together with some natural facts, metaphysically necessitate moral facts – so there exists a metaphysical necessitation relation between natural facts and moral ones. And, on **Modified Difference-Making**, things that stand in a metaphysical necessitation relation to a cause are exempted from being held fixed when testing for causation.

With these considerations in mind, we can again plug cases into **Modified Difference-Making** to see how non-naturalist moral causation might go. Consider again my friend breaking her promise. When testing for causation, we should not consider the metaphysically impossible world – and it is metaphysically impossible, since moral laws are metaphysically necessary – where certain natural facts (that my friend drove her car past the airport without stopping, and so on) are the same but there is no moral fact, to see if I still would have thought that something bad went on. We should instead consider the metaphysically possible world where none of the natural facts or the moral facts obtain. And in such a world, presumably, I would not have thought that something bad went on. So the moral fact that it was wrong for my friend to break her promise caused my thinking that something bad went on. Again, *voilà* – moral causation on non-naturalism.

Such a claim has precedent: Woodward (2015) makes a similar move in regard to metaphysically necessary laws in nonreductive physicalist ontologies, and White (2018) and Brown (2019) do in regard to metaphysically necessary psychophysical laws in dualist ontologies, in order to posit mental causation.

So, even better – this moral causation on non-naturalism is not motivated by concerns in metaethics. It is motivated by concerns in the causation literature about accounting for higher-level causation. It simply turns out that when we modify our frameworks for causation to account for these things, we get moral causation as well. The non-naturalist should therefore feel confident that this is a philosophically defensible framework for non-naturalist moral causation.

4.

Armed with **Modified Difference-Making**, we may choose which premise of the exclusion argument we want to deny.

I think the happiest option is denying **Exclusion**. But it is not clear we may do so – it seems plausible that moral facts, despite being causally efficacious, overdetermine their effects in conjunction with their underlying natural phenomena on standard tests for overdetermination (Bennett 2003, 2008). To argue that these facts don't overdetermine, we'll at least have to say more about what non-naturalist ontologies tell us about nearby possible worlds in which there exists my friend's promise-breaking but not her driving by the airport, and *vice versa* – and I have no space to do that here.

We may more easily deny **Causal Closure**, just by introducing a proportionality requirement (Yablo 1992, 1997, Strevens 2008, List and Menzies 2009, Raatikainen 2010, Zhong 2020a,b, Vaassen 2022, forthcoming, Rubenstein forthcoming) to **Modified Difference-Making**. With the introduction of such a requirement, moral causes will outcompete natural causes (Oddie 2005). My friend's acting wrongly causes my feeling that something bad went on because, e.g., the natural fact that my friend drove her car past the airport without stopping is too general to be a cause (she could have done this without making me feel that something bad was going on, if she called me to tell me she was rushing to the bedside of a sick relative), and e.g., the natural fact that all relevant particles were arranged in a certain manner is too specific (I would have still felt something bad was going on if she had one fewer electron).

If we want to keep **Causal Closure**, we may deny **Non-Overdetermination**. Then we will say that moral facts cause natural effects, and these effects also are caused by the underlying natural phenomena of the moral facts (such as my friend driving her car past the airport). Pereboom (2002), Sider (2003), Schaffer (2003), and Loewer (2007b) adopt this strategy in the mental causation debate, accepting overdetermination to secure mental causation without giving up physical completeness.

We may pick whichever of these options seems least objectionable. On any of these, however, our results are theoretically well-motivated by considerations in the causation literature, again giving us confidence that this is a plausible account of non-naturalist moral causation.

5.

There are two upshots I want to bring out here.

The first has to do with the naturalism/non-naturalism dialectic.

Moral causation has been upheld as a potential dialectical advantage for naturalists (Majors 2003, Cuneo 2006, McBrayer 2010, Enoch 2017, Stringer 2021). Look, the naturalist might say to the non-naturalist, I can account for this thing you can't account for, so naturalism is the better theory. But it is not better on these grounds, as non-naturalists may posit moral causation after all.

The second has to do with non-naturalist moral epistemology.

There is a family of skeptical challenges to non-naturalism that go: if moral facts are non-natural, and we have reliable epistemic access to these facts, how in the world do we have this access (Enoch, 2011, Schechter, 2018)? Non-naturalists generally put forward non-causal explanations for this (e.g., Enoch, 2011, Wielenberg, 2014, Baron et al., 2019).

There is a glass-half-empty thing to say here, which is: if I am right, such non-naturalist moral epistemology simply rests on a mistake. Non-naturalist moral epistemologists are trying to explain how we have reliable epistemic access to non-causally-efficacious non-natural moral facts. But non-naturalists needn't posit such non-causally-efficacious facts, nor should they. So nothing we have said on this topic so far is of any use.

There is also a glass-half-full interpretation for the non-naturalist, which is: non-naturalists shouldn't worry about this challenge, as they can easily explain our epistemic access via moral causation. So again an advantage for naturalists disappears — naturalists and non-naturalists alike can claim explainable epistemic access to the moral.

This is to say: the possibility of non-naturalist moral causation is interesting in its own right. But it should be even more interesting to the non-naturalist who wants to retake some dialectical territory from the naturalist. Such a non-naturalist, then, ought to be happy to hear that our best accounts of causation deliver non-naturalist moral causation without much trouble at all.

References

- Albert, David Z. (2015) *After Physics*, Harvard University Press.
- Audi, Robert (1993) "Mental Causation: Sustaining and Dynamic", in: John Heil and Alfred R. Mele, editors, *Mental Causation*. Oxford University Press.
- Audi, Robert (1997) "Ethical Naturalism and the Explanatory Power of Moral Concepts", in: *Moral Knowledge and Ethical Character*, Oxford University Press.
- Baron, S.; Colyvan, M.; Miller, K.; and Rubin, M. (2019) "Non-Naturalistic Moral Explanation", *Synthese*, 198(5), 4273–4294.
- Baumgartner, Michael (2010) "Interventionism and Epiphenomenalism", *Canadian Journal of Philosophy*, 40(3), 359–383.
- Bennett, Jonathan (2003) *A Philosophical Guide to Conditionals*, Oxford University Press.
- Bennett, Karen (2003) "Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It", *Noûs*, 37(3), 471–97.
- Bennett, Karen (2008) "Exclusion Again", in: Jakob Hohwy and Jesper Kallestrup, editors, *Being Reduced: New Essays on Reduction, Explanation, and Causation*, 280–307, Oxford University Press.
- Berker, Selim (2018) "The Explanatory Ambitions of Moral Principles", *Noûs*, 53(4), 904–936.
- Bhagal, Harjit (forthcoming) "Moral Necessitism and Scientific Contingentism", *Oxford Studies in Metaethics*.
- Blanchard, Thomas (2020) "Explanatory Abstraction and the Goldilocks Problem: Interventionism Gets Things Just Right", *British Journal for the Philosophy of Science*, 71(2), 633–663.
- Brown, Christopher Devlin (2019) "Exclusion Endures: How Compatibilism Allows Dualists to Bypass the Causal Closure Argument", *Analysis*, 79(4), 587–594.
- Campbell, John (2010) "Control Variables and Mental Causation", *Proceedings of the Aristotelian Society*, 110(1pt1), 15–30.

- Copp, David (2007) *Morality in a Natural World: Selected Essays in Metaethics*, Cambridge Studies in Philosophy, Cambridge University Press.
- Cuneo, Terence (2006) "Moral Facts as Configuring Causes", *Pacific Philosophical Quarterly*, 87(2), 141–162.
- Darwall, S.; Gibbard, A.; and Railton, P. (1992) "Toward Fin de siècle Ethics: Some Trends", *The Philosophical Review*, 101(1), 115–189.
- Dworkin, Ronald (1996) "Objectivity and Truth: You'd Better Believe It", *Philosophy & Public Affairs*, 25(2), 87–139.
- Enoch, David (2011) *Taking Morality Seriously: A Defense of Robust Realism*, Oxford University Press.
- Enoch, David (2017) "Non-Naturalistic Realism in Metaethics", in: Tristram McPherson and David Plunkett, editors, *The Routledge Handbook of Metaethics*, 29– 42, Routledge.
- Enoch, David (2019) "How Principles Ground", *Oxford Studies in Metaethics*, 14, 1–22.
- Fogal, D. and Risberg, O. (2020) "The Metaphysics of Moral Explanations", in: Russ Shafer-Landau, editor, *Oxford Studies in Metaethics Volume 15*, 170–194, Oxford University Press.
- Goodman, N. (1973) *Fact, Fiction, and Forecast*, Harvard University Press.
- Harman, G. (1977) *The Nature of Morality: An Introduction to Ethics*, Oxford University Press.
- Ismael, J. (2016) *How Physics Makes Us Free*, Oxford University Press.
- Kim, Jaegwon (1989) "The Myth of Non-Reductive Materialism", *Proceedings and Addresses of the American Philosophical Association*, 63(3), 31–47.
- Kim, Jaegwon (1998) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, MIT Press.
- Kim, Jaegwon (2005) *Physicalism, or Something Near Enough*, Princeton University Press.
- Leary, Stephanie (2017) "Non-Naturalism and Normative Necessities", *Oxford Studies in Metaethics*, 12.
- Lewis, David (1973) "Causation" *Journal of Philosophy*, 70(17), 556–567.
- Lewis, David (1979) "Counterfactual Dependence and Time's Arrow" *Noûs*, 13(4), 455–476.
- List, Christian and Menzies, Peter (2009). "Nonreductive Physicalism and the Limits of the Exclusion Principle", *Journal of Philosophy*, 106(9), 475–502.

- Loewer, Barry (2007a) "Counterfactuals and the Second Law", in: Huw Price and Richard Corry, editors, *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford University Press.
- Loewer, Barry (2007b) "Mental Causation, or Something Near Enough", in: Brian P. McLaughlin and Jonathan D. Cohen, editors, *Contemporary Debates in Philosophy of Mind*, 243–64, Blackwell.
- Majors, B. (2003) "Moral Explanation and the Special Sciences", *Philosophical Studies*, 113(2), 121–152, 2003.
- Malcolm, Norman (1968) "The Conceivability of Mechanism", *Philosophical Review*, 77 (January), 45–72.
- McBrayer, Justin P. (2010) "Moral Perception and the Causal Objection", *Ratio*, 23(3), 291–307.
- Colin McGinn (1997) *Ethics, Evil, and Fiction*, Oxford University Press.
- McPherson, Tristram and Plunkett, David (2023). "Ground, Essence, and the Metaphysics of Metanormative Non-Naturalism", *Ergo*, 9.
- Oddie, Graham (2005) *Value, Reality, and Desire*, Oxford University Press.
- Papineau, David (2002) *Thinking About Consciousness*, Oxford University Press.
- Papineau, David (2013) "Causation is Macroscopic but Not Irreducible", in: S. C. Gibb and Rognvaldur Ingthorsson, editors, *Mental Causation and Ontology*, Oxford University Press.
- Papineau, David (2022) "The Statistical Nature of Causation", *The Monist*, 105(2), 247– 275.
- Pereboom, Derk (2002) "Robust Nonreductive Materialism", *The Journal of Philosophy*, 99(10), 499–531.
- Raatikainen, Panu (2010) "Causation, Exclusion, and the Special Sciences", *Erkenntnis*, 73(3), 349–363.
- Rosen, Gideon (2017a) "Metaphysical Relations in Metaethics", in: Tristram McPherson and David Plunkett (eds), *The Routledge Handbook of Metaethics*, 151– 169, Routledge.
- Rosen, Gideon (2017b) "Ground by Law", *Philosophical Issues*, 27(1), 279–301.
- Rosen, Gideon (2017c) "What is a Moral Law?", *Oxford Studies in Metaethics*, 12.
- Rubenstein, Ezra (forthcoming) "Cohesive Proportionality", *Philosophical Studies*, 1–25.
- Scanlon, Thomas (2014) *Being Realistic About Reasons*, Oxford University Press.
- Schaffer, Jonathan (2003) "Overdetermining Causes", *Philosophical Studies*, 114(1-2), 23–45.

Schechter, Joshua (2018) "Explanatory Challenges in Metaethics", in: Tristram McPherson and David Plunkett (eds), *Routledge Handbook of Metaethics*, 443–459, Routledge.

Shafer-Landau, Russ (2003) *Moral Realism: A Defence*, Oxford University Press.

Shapiro, Larry and Sober, Elliott (2007) "Epiphenomenalism - the Do's and the Don't's", in: G. Wolters and Peter K. Machamer (eds), *Thinking About Causes: From Greek Philosophy to Modern Physics*, 235–264, University of Pittsburgh Press.

Sider, Theodore (2003) "What's so Bad about Overdetermination?", *Philosophy and Phenomenological Research*, 67(3), 719–726.

Skiles, Alexander (2015) "Against Grounding Necessitarianism", *Erkenntnis*, 80(4), 717– 751.

Slutsky, David (2001) "Causally Inefficacious Moral Properties", *Southern Journal of Philosophy*, 39(4), 595–610.

Smith, Michael (1994) *The Moral Problem*, Blackwell, Cambridge, Mass., USA.

Strevens, Michael (2008) *Depth: An Account of Scientific Explanation*, Harvard University Press.

Ryan Stringer, Ryan (2021) "Ethical Emergentism and Moral Causation", *Journal of Moral Philosophy*, 18(4), 331–362.

Thomson, Judith Jarvis (1996) "Epistemological Arguments for Moral Scepticism", in: Gilbert Harman and Judith Jarvis Thomson (eds), *Moral Relativism and Moral Objectivity*, Blackwell.

Vaassen, Bram (2022) "Halfway Proportionality", *Philosophical Studies*, (9):1–21.

Vaassen, Bram (forthcoming) "Mental Causation for Standard Dualists", *Australasian Journal of Philosophy*.

Viggiano, Andrea (2020) "Consequentialism and the Causal Efficacy of the Moral", *Philosophical Studies*, 177(10), 2927–2944.

White, Ben (2018) "Metaphysical Necessity Dualism", *Synthese*, 195(4), 1779–1798.

Wielenberg, Erik (2014) *Robust Ethics*, Oxford University Press.

Woodward, James (2003) *Making Things Happen: A Theory of Causal Explanation*, Oxford Studies in Philosophy of Science. Oxford University Press, Oxford, New York.

Woodward, James (2008) "Mental Causation and Neural Mechanisms", in: Jakob Hohwy and Jesper Kallestrup, editors, *Being Reduced: New Essays on Reduction, Explanation, and Causation*, 218–262. Oxford University Press.

Woodward, James (2014) "A Functional Account of Causation; or, a Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters – Usefulness", *Philosophy of Science*, 81(5), 691–713.

Woodward, James (2015) "Interventionism and Causal Exclusion", *Philosophy and Phenomenological Research*, 91(2), 303–347.

Woodward, James (2020) "Causal Complexity, Conditional Independence, and Downward Causation", *Philosophy of Science*, 87(5), 857–867.

Yablo, Stephen (1992) "Mental Causation", *The Philosophical Review*, 101(2), 245–280.

Yablo, Stephen (1997) "Wide Causation" *Noûs*, 31(s11), 251–281.

Zhong, Lei (2020a) "Intervention, Fixation, and Supervenient Causation", *Journal of Philosophy*, 117(6), 293–314.

Zhong, Lei (2020b) "Taking Emergentism Seriously", *Australasian Journal of Philosophy*, 98(1), 31–46.

Acknowledgement

Thanks to Karen Bennett, Andy Egan, Carolina Sartorio, Jonathan Schaffer, Bram Vaassen, and Itamar Weinshtock Saadon for their helpful comments on various drafts of this paper.

Der *Tractatus Logico-Philosophicus* und die Rede vom Verstehen

Moritz René Pretzsch (Kassel, Germany)

Abstract

Untersuchungsgegenstand dieses Beitrages ist Ludwig Wittgensteins Rede vom „Verstehen“ im *Tractatus Logico-Philosophicus*, die sich wie ein roter Faden durch seine Abhandlung hindurchzieht. Auffällig ist, dass der Begriff des Verstehens im Vorwort und in Satz 6.54 des *Tractatus* von Wittgenstein anders verwendet wird als in den mittleren Sätzen. So geht der Text an diesen Stellen in eine Selbstreflexion und möchte etwas über sich selbst aussagen. Weiter von Bedeutung ist in diesem Zusammenhang, dass das Verstehen in einer wichtigen Beziehung zu dem Begriff des Erläuterns steht. Wittgenstein möchte dem Leser seinen *Tractatus* verständlich machen, indem er diesem seine Sätze primär erläutert. Erklärungen scheinen dagegen Mittel zum Zweck zu sein und auf eine Belehrung möchte Wittgenstein bekanntlich gänzlich verzichten. Für ihn muss ein philosophisches Werk, das der Leser verstehen möchte, wesentlich aus Erläuterungen bestehen. In diesem Beitrag möchte ich dafür argumentieren, dass Wittgenstein den Leser qua eines gedanklich-handelnden Vollzuges zu einem *richtigen* Verstehen führen möchte, im Sinne einer Läuterung. Dabei werde ich die verschiedenen Arten vom Verstehen, welche im *Tractatus* bestehen, näher beleuchten und die wichtige Beziehung zwischen Verstehen und Erläutern aufzeigen.

Im *Tractatus logico-philosophicus* ist häufig die Rede von „auffassen“, „erkennen“, „ersehen“, „sehen“, „richtig sehen“ und „klar werden“. Insbesondere aber ist die Rede von „verstehen“. Richtet man den Blick gezielt darauf, so zeigt sich, dass Wittgenstein sowohl auf den mittleren Sprossen der Leiter als auch am Anfang und Ende des *Tractatus* vom „verstehen“ spricht. Auffällig jedoch ist, dass der Begriff des Verstehens im Vorwort des *Tractatus* und in Satz 6.54 von Wittgenstein anders verwendet wird, als in den mittleren Sätzen. Ist an diesen Stellen von „verstehen“ die Rede (vgl. TLP, 3.334, 4.003–4.1213, 4.4243, 4.411, 5.02, 5.451, 5.521–5.62, 6.2341), so geht dies meist mit einer rein pragmatischen Verwendung des Begriffs einher, verbunden mit einem technischen Zweck. Dient die Verwendung, innerhalb der Hauptsätze drei bis sechs, doch vornehmlich dazu, den Akt der Projektion von logischen Bildern (Sätzen, Gedanken) auf die Sachverhalte der Welt zu benennen. So begegnet man dem Verstehen etwa an den zentralen Strukturelementen der Abbildtheorie: Einen Satz zu verstehen (vgl. TLP, 4.021), heißt, wissen, was der Fall ist (die Tatsachen kennen), wenn er wahr ist; einen Namen verstehen (vgl. TLP, 4.243), heißt, den von ihm bezeichneten Gegenstand kennen; und die logischen Verknüpfungen letztlich müssen „sich von selbst verstehen“ – wie es die Grundidee des *Tractatus* – dass die Logik für sich selber sorgen müsse –

einfordert (vgl. TLP, 3.334, 5.473). Ganz anders gelagert scheint die Verwendung des Begriffs Verstehen dagegen im Vorwort und im letzten Satz der sechsten Leitersprosse, Satz 6.54. Welche im Folgenden näher untersucht werden soll. Im Vorwort heißt es:

Dieses Buch wird vielleicht nur der verstehen, der die Gedanken, die darin ausgedrückt sind – oder doch ähnliche Gedanken – schon selbst einmal gedacht hat. – Es ist also kein Lehrbuch. – Sein Zweck wäre erreicht, wenn es Einem, der es mit Verständnis liest, Vergnügen bereitet. Das Buch behandelt die philosophischen Probleme und zeigt – wie ich glaube – dass die Fragestellung dieser Probleme auf dem Mißverständnis der Logik unserer Sprache beruht. (TLP, *Vorwort*)

Eines steht hier fest, mit Blick auf die Bemerkung zum Verstehen im Vorwort: Was auch immer die individuelle Leseerfahrung als Verständnis für sich reklamieren mag und wie groß dabei das „Vergnügen“ (vgl. TLP, *Vorwort*) sein mag – das Verständnis kann weder in der Kenntnis neuer Sachverhalte bestehen noch von selbst eintreten. Es bedarf der von Wittgenstein verfertigten Gedanken, wenngleich diese nicht als „Lehre“ aufzufassen sind. Und doch stellt sich zwangsläufig die Frage, welchen Zugewinn *der* Lesende erhält, der die „die Gedanken“, die im *Tractatus* ausgedrückt wurden, „schon selbst einmal gedacht hat.“ (TLP, *Vorwort*) Wenngleich Wittgenstein noch eine Relativierung einschiebt: „oder doch ähnliche Gedanken“ (TLP, *Vorwort*). Es scheint nicht leicht, darüber Klarheit zu gewinnen, denn einerseits setzt Wittgenstein ein vorläufiges Denken, der im *Tractatus* selbst aufgestellten Gedanken voraus – was ein massives Einschränken der Zielgruppe bedeutet – andererseits aber, weitet Wittgenstein die Einschränkung der Zielgruppe qua Relativierung etwas aus. Zu konstatieren ist hierbei sicherlich, dass Wittgenstein ein *vorläufiges* Denken voraussetzt, nicht aber ein vorläufiges Verstehen. Wittgenstein möchte dem Lesenden ein bestmögliches Verstehen ermöglichen, indem er versucht, die Gedanken so gut wie nur möglich auszudrücken, den Nagel bestmöglich auf den Kopf zu treffen und ein „Rauschen und Brausen“ zu vermeiden (vgl. TLP, *Vorwort* und *Motto*). Er trifft hierbei eine Aussage über das gesamte Verstehen seines Buches selbst und möchte den Leser auf den Charakter der nachfolgenden Sätze einstimmen.

In Satz 6.54 heißt es:

Meine Sätze erläutern dadurch, dass sie der, welcher mich versteht, am Ende als unsinnig erkennt, wenn er durch sie – auf ihnen – über sie hinausgestiegen ist. (Er muss sozusagen die Leiter wegwerfen, nachdem er auf ihr hinaufgestiegen ist.) Er muss die Sätze überwinden, dann sieht er die Welt richtig. (TLP, 6.54)

In Satz 6.54 setzt eine Rückschau ein und es wird die Wirkung umrissen, die das Erklimmen der Sprossen beim Leser evoziert haben soll. Wittgenstein wendet sich hierbei erneut an den Leser und bezieht sich auch hier auf ein Gesamtverständnis des *Tractatus*. Wenn Wittgenstein hier vom Verstehen spricht, so ist dies kein Verstehen, das sich nahtlos in die Denksituation des Haupttextes einzufügen weiß. Und erst *diese* Rede vom Verstehen, wie sie in Satz 6.54 geäußert wird, lässt den Lesenden den Autor selbst begegnen. Diese Rede vom Verstehen zeigt: So zeitlos und apodiktisch der Mehrzahl der nummerierten Sätze formuliert sein mag, und so gleichsam selbstgenügsam der Text seinem Leser die meiste Zeit gegenübertritt – als Ganzes strebt ebendieser Text zuerst danach, von *jemandem* verstanden zu werden. Mehr noch: Der Text, der sich an dieser Stelle selbst mit einer Leiter vergleicht, scheint letztlich eine Leiter zur Person des Autors sein zu wollen. Dabei lässt sich auch eine feine Diskrepanz bei der Adressierung herauslesen: Heißt es am Fuße der Leiter, im Vorwort, noch: „derjenige“ der das Buch versteht und „die Anderen“, die „ähnliche Gedanken“ schon gedacht haben (vgl. TLP, *Vorwort*). So wendet sich Wittgenstein am Ende der Leiter dann demjenigen zu, „welcher mich versteht“ (TLP, 6.54).

Insgesamt lässt sich feststellen: Sowohl im Vorwort als auch innerhalb des Satz 6.54 wird der Begriff des Verstehens von Wittgenstein auffällig weit gefasst, und es lässt sich beobachten, dass an diesen Stellen der Text in Selbstreflexion geht, und etwas über sich selbst aussagen möchte. Weiter zeigt sich: Am Fuße der Leiter erhält der Leser den Hinweis, dass sie ihn zu einem allgemeinen Verstehen führen soll; auf den Leitersprossen wird ihm dann ein spezifisches, aber auch sehr präzises Verständnis davon vermittelt, was verstehen eigentlich heißt; am Zielende der Leiter erfährt er dann, dass er sich bei jenem präzisen Verstehen nicht beruhigen darf, sondern dass ihn die Leiter letztlich zu einem *anderen* Verstehen hinführen möchte. Der Leser soll das Buch verstehen, dieses lehrt ihn, die Sprache zu verstehen – indem er am Ende die „Unsinnigkeit“ erkennt und gleichfalls erkennt, „wie wenig damit getan ist“,

verstehen er dessen Autor. Nimmt man das Bild der Leiter ernst, so liegt es nahe, sich das Verstehen als einen der beiden Leiterholme vorzustellen, denn es begleitet den Leser als Hoffnung und Aufforderung des Autors durch die gesamte Aufstiegsbewegung des *Tractatus*. Ebenso ist das Verstehen die Aktivität, die von dem Leser durchgehend versucht wird. Das scheint *prima facie* eine Trivialität zu sein – wie soll denn einer ein Buch lesen, ohne durchgehend verstehen zu wollen? –, doch im Fall des *Tractatus* ist es das nicht. Den *Tractatus*, im sozusagen höheren Sinn, zu verstehen, fällt ja laut 6.54 mit der Erkenntnis zusammen, seine Sätze seien unsinnig – also gerade *nicht* verständlich. Dieses Paradoxon löst sich erst auf, wenn der Leser von einer theoretischen zu einer performativen Lesehaltung findet. Innerhalb des *Tractatus* begegnet der Leser Worten und Begriffen, deren Bedeutung nicht fixiert sind, auch wenn die strenge Art ihrer Einführung dies zunächst erwarten lässt. Es gilt, dass sich der Leser darauf offen einlässt; an der Bedeutung so lange festhält, bis sie der Text selbst erweitert oder korrigiert. Erst im Rückblick kann dem Leser klar werden, *was* sich in dem Bedeutungswandel durchgehalten hat – weshalb es gerade dieser Begriff war, der sich auf einer höheren Sprosse wandeln musste, und weshalb es gerade jener Begriff war, den er unverändert, unbekümmert auf einer Sprosse zurücklassen konnte. All dies weiß sich durch die Schwierigkeit ergänzt, dass sich die Begriffe diesen Wandel bekanntlich nicht anmerken lassen, während er sich vollzieht. Der Text *ist* die Leiter. Der Leser muss die Bedeutungen zunächst festhalten, um etwas verstehen zu können – genauso, wie er mit seinen Sätzen zunächst umgehen müssen, als hätten sie Sinn. Die Metapher der Leiter ernst zu nehmen, heißt den Gedankengang ernst zu nehmen und sich tatsächlich auf den Weg zu machen. Zum richtigen Verständnis muss letztlich gehören, dass die Sätze als unsinnig erkannt werden, wie es ebenso wesentlich dazugehört, sie erst *am Ende* als unsinnig zu erkennen. Der Leser muss durch ein verengtes Verstehen hindurch, um den Punkt zu erreichen, wo ein umfassendes Verstehen möglich ist. Erst am abschließenden Aussichtspunkt hat der Leser wirklich verstanden, nämlich nicht bloß das Buch, sondern auch seinen Autor.

Wenn die hier angeführten Bemerkungen zum Verstehen Wittgensteins Anliegen gerecht wurden, so wird auch klar, weshalb der *Tractatus* selbst auf explizite Erläuterungen, wie sie hier gegeben werden, verzichten muss: Es geht

um einen *Wechsel* der Art, *wie* der Leser versteht. Den Modus des Verstehens kann er jedoch nur in einem gedanklichen Handeln verändern, nicht indem er im Ausgangsmodus verbleibt und in diesem immer *mehr* versteht. Eine nebenherlaufende Meta-Erzählung, die den Leser immer wieder darauf hinweist, dass er jetzt anders verstehen soll, könnte dieser ja doch nur wieder *so* verstehen, wie er sie *bereits* versteht. Das Verstehen von Sätzen und Argumenten lässt sich nicht dadurch zum Verstehen des Gegenübers, der Person erheben, indem man diesem immer weitere Sätze und Argumente vorlegt. Das Verstehen muss sich damit selbst überwinden. Der *Tractatus* lässt sich damit als eine „Anleitung“ zu exakt diesem gedanklich-handelnden Vollzug verstehen. Dabei wird der gedanklich-handelnde Vollzug von Wittgenstein *nicht* explizit gemacht, sondern er findet sich *implizit* angelegt. Wittgenstein möchte den Leser meines Erachtens nicht durch Erklärungen, sondern v. a. durch Erläuterungen zu einem richtigen Verstehen führen. Erklärungen sind lediglich Mittel zum Zweck. Wittgenstein wählt das Verb *erläutern*, um das eigentümliche Vorgehen zu beschreiben, mit dem der *Tractatus* sein Ziel erreichen möchte: dass nämlich *er*, Wittgenstein, verstanden wird. Das Kerngeschäft des *Tractatus* besteht demnach im Erläutern – was zunächst an das Vorwort erinnert, wo der Charakter des Buches dezidiert nicht, den eines Lehrbuches haben soll. In einem solchen erwartete man wohl in erster Linie Lehrsätze, Beweise oder Definitionen, vielleicht Übersichten; jedenfalls weniger und allenfalls in zweiter Linie Erläuterungen. Dazu passt die häufig zu machende Beobachtung, dass Lehrbücher im engeren Sinn alle erläuterungsartigen Zusätze (wie Beispiele, Anwendungen oder historische Hintergründe) gern in einen Kasten auslagern, der vom übrigen Text in ähnlicher Weise abgesetzt ist wie Bilder und Diagramme, die ja ebenfalls Erläuterungscharakter besitzen. Die Bedeutung des Erläuterns kann meines Erachtens nicht hoch genug eingeschätzt werden, wenn es um ein richtiges Verstehen des *Tractatus* geht. Für Wittgenstein besteht ein „philosophisches Werk [...] wesentlich aus Erläuterungen.“ (TLP, 4.112) Warum gerade das Erläutern für das Verstehen des *Tractatus* wichtig ist – in der einschlägigen Sekundärliteratur bislang kaum untersucht (vgl. etwa Apel 1966; Gruender 1962; Venieri 1989; Ben-Menahem 1998; Gunnarsson 2000; Tatjevskaya 2013) –, und nicht primär das Erklären, bildet Inhalt des kommenden Abschnitts.

Begonnen sei mit einem Blick auf die semantische Ebene: Die Basis sowohl von *erläutern* wie von *erklären* ist jeweils ein Adjektiv: *lauter* im ersten, *klar* im zweiten Fall. Bereits diese Basen lassen eine starke semantische Ähnlichkeit erkennen, ebenso wie einen spezifischen semantischen Unterschied, der sich bis auf die Endstufe unserer Verben erhalten wird. *Lauter* und *klar* haben beide zunächst eine handfeste Bedeutung: in etwa „rein“ oder „sauber“. *Lauter* ist seit dem Althochdeutschen belegt und hat seinen Ursprung in einem konstruierten indogermanischen Verb mit der Bedeutung „waschen“ oder „spülen“; während *klar* eine mittelhochdeutsche Entlehnung aus dem Französischen ist und aufs lateinische *clarus*, also „hell“, zurückgeht (vgl. Kluge 2012). Während *lauter* vorzugsweise auf den seelisch-moralischen Bereich übertragen wird – indem wir etwa von *lauteren Absichten* oder einem *lauteren Herzen* sprechen –, wird *klar* umfassender für den gesamten abstrakt-intellektuellen Bereich verwendet – so sind *klare Aussagen* besonders deutlich und präzise getroffene Aussagen, eine *klare Argumentation* ist übersichtlich und von Überflüssigem und Ungenauem gereinigt. Eine Wertung über die jeweilige moralische Qualität ist hier jedoch nicht mehr impliziert.

In einem zweiten Schritt bildet das Deutsche aus beiden Adjektiven zwei entsprechende faktitive Verben. Die etwas grobschlächtige Bedeutungsparaphrase für faktitive Verben lautet „x machen“ und in dem hier vorliegenden Fall wird diese neue Verwendungsweise der ursprünglichen Adjektive durch den Umlaut sowie die kontextabhängigen Verbendungen angezeigt. Entscheidend ist: Zum einen erfordert die syntaktische Struktur von *läutern* und *klären* notwendig zwei Beteiligte: Typischerweise einen Akteur und einen Gegenstand, nämlich erstens jemanden, der zweitens etwas läutert oder klärt. Zum anderen hat sich die semantische Differenz von *lauter* und *klar* auf dieser zweiten Ebene der faktitiven Verben erhalten. Es gibt für beide Verben handfeste Verwendungsweisen. Ihre Wahl scheint nicht völlig gleichgültig.

Im dritten und letzten morphologischen Schritt werden die beiden soeben betrachteten Verben um das Präfix *er-* erweitert. Strukturell bewirkt die Präfigierung in beiden Fällen eine Erweiterung des Verbs um einen weiteren syntaktischen Mitspieler. Die Standardfälle von *läutern* und *klären* weisen ein Subjekt und ein Akkusativobjekt auf. Inhaltlich erhält sich zwar eine gewisse Differenz, verglichen mit den vorigen Stufen ist diese jedoch stärker verwischt.

Und dennoch: Ein aufmerksamer Sprachsinn, jenseits des Alltagsrauschens, hört immer noch die jeweilige Herkunft der Wörter heraus. So bewirkt die Verwurzelung von *lauter* in der seelisch-moralischen Sphäre eine etwas freiere, gleichsam pädagogische Note im Verb *erläutern*; wohingegen dem Verb *erklären*, entsprechend seiner hier beleuchteten abstrakt-intellektuellen Herkunft, eine striktere, gleichsam wissenschaftliche Note eignet. Dieses zweite Verb spielt zudem eine *wichtige* Rolle in den späteren Schriften Wittgensteins, wo es zumeist zur Charakterisierung dessen dient, worum es ihm und seinem Philosophieren *nicht* geht (vgl. PU, § 126 und BBB, 39). Ohne dies im Folgenden weiter zu berücksichtigen, deutet sich an, dass diese kleine, scheinbar harmlose Betrachtung dieses Beitrags möglicherweise an einem Kernbestandteil von Wittgensteins Denken rührt – und damit sogar für das Verständnis der späteren Schriften fruchtbar gemacht werden könnte.

Erläuterungen scheinen tendenziell auf (gemeinsames) Verstehen zu zielen, Erklärungen dagegen auf (einseitige) Wissensvermittlung. Wie weiter gezeigt werden konnte, ist im *Erläutern* aufgrund seiner Wortgeschichte und Wortbauweise eine moralische Tiefendimension enthalten – vieles spricht dafür, dass es in Wittgensteins Interesse ist, sie mitzuhören. Was folgt nun darauf für das Verstehen des *Tractatus*?

Zunächst sei der Blick erneut auf Satz 6.54 gerichtet: Dass in 6.54 das Verb *erläutern* – und nicht etwa *erklären* oder *nachweisen* – steht, kann meines Erachtens schwerlich als stilistischer Zufall abgetan werden; vielmehr scheint es der präzise und angemessene Ausdruck für dasjenige zu sein, was der *Tractatus* am Leser vollführen möchte. Er will ihm keine Lehrsätze vermitteln und ihn auch nichts inhaltlich Neues lehren, sondern seinen Blick klären oder neu ausrichten, wozu es weniger einer strengen Argumentation als einer behutsamen philosophischen Pädagogik bedarf. Wie so oft in der Philosophie weiß sich auch hier ein Primat des Wie gegenüber einem Was situiert: „Das Resultat der Philosophie sind nicht ‚philosophische Sätze‘, sondern das Klarwerden von Sätzen“ (TLP, 4.112), schreibt Wittgenstein. Damit kommt es gewissermaßen zu einer Zurückweisung einer gehaltvollen philosophischen Erkenntnis. Es gibt keine substanziellen philosophischen Gehalte und auch keinen besonderen Gegenstandsbereich der Philosophie – dergleichen kommt nur den Naturwissenschaften zu. Aber es gibt einen besonderen Modus, der zwar auch nicht die Philosophie selbst ist, zu dem diese jedoch *erläuternd*

hinführen kann; er *zeigt* sich dort, wo die Probleme verschwinden und die Sachen einfach und klar liegen. Das Ende von 6.54 bestätigt es nochmals: Wer versteht, *sieht* die Welt richtig. Im Verb erläutern findet nun all dies seinen Ausdruck, ja in ihm klingt sogar jene Wende ins Ethisch-Moralische an, die sich gegen Ende des *Tractatus* abzeichnet und die dem Leser den einzigen Hinweis gibt, an welchem Ort er sich eigentlich wiederfinden wird, nachdem wir die Leiter emporgeklommen ist. Kurz, den *Tractatus* als eine Art von Anleitung zu einem bestimmten gedanklichen Vollzug zu lesen, findet im „erläutern“ von Satz 6.54 eine wichtige Stütze. Der gedankliche Vollzug, wie er gerade auch bei der therapeutischen Lesart des *Tractatus* von zentraler Bedeutung ist (vgl. Winch 1969; Conant 1990; Diamond 1991), steht dabei letztlich im Mittelpunkt.

Schlussbemerkung:

Wittgenstein hofft auf eine Einsicht seitens des Lesers. Eine Einsicht, die das Leben des Leser selbst betrifft und den richtigen Blick auf die Welt eröffnet. Um den richtigen Blick auf die Welt zu gewinnen, kann die Philosophie von Nutzen sein, aber im Erreichen ihres Zieles verzehrt sie sich selbst. Die Einsicht, zu der die Philosophie führt, geht Hand in Hand mit dem Verlust des philosophischen Bedürfnisses. Die Leiter benötigt nur derjenige, der noch nicht oben angekommen ist; dann allerdings muss er sie wegwerfen. Das philosophische Ziel ist im eminenten Sinn die Überwindung der Philosophie, und wenn die Philosophie wiederum eine Tätigkeit des Einzelnen ist, dann ist ihr Resultat zuallererst eine Form der Selbstüberwindung: Der Leser hat sein eigenes Bedürfnis, sein Verlangen nach Philosophie überwunden. Diese Überlegungen erlauben es, die Einsicht zu formulieren, dass jene *Erläuterung*, die Wittgensteins Sätze geben wollen, in einem wichtigen Sinn eine *Selbstläuterung* bedeuten. Der *richtige* Blick, zu dem der *Tractatus* verhelfen möchte, ist nicht einfach bloß ein klarer Blick – es ist der *geläuterte* Blick auf die Welt. An ihrem Bestand hat sich nichts geändert, der Leser hat nicht einmal etwas Neues in ihr zu entdecken; verändert worden ist einzig der Modus, der Blick des Lesers auf die Welt. Der Leser versteht sein anfängliches Verstehen anders und darum gibt es letztlich auch nichts, das im eigentlichen Sinn erläutert wurde, sondern was bleibt ist eine reine Erläuterung, die alles umfasst und

dadurch Selbstzweckcharakter erhält. Was der Leser sieht, bleibt gleich; *wie* er es sieht, das hat sich geändert. Erst dann hat der Leser *verstanden*, kann die Leiter wegwerfen und seinen *geläuterten* Blick im Handeln beweisen.

References

Apel, Karl Otto (1966) „Wittgenstein und das Problem des hermeneutischen Verstehens“, *Zeitschrift für Theologie und Kirche* 63.1, 49 – 87.

Ben-Menahem, Yemima (1998) “Explanation and description: Wittgenstein on convention”, *Synthese* 115.1, 99 – 130.

Conant, James (1990) “Throwing Away the Top of the Ladder”, *Yale Review* 79, 328 – 364.

Diamond, Cora (1991) *The realistic spirit: Wittgenstein. Philosophy, and the Mind*, Boston: MIT Press.

Gruender, David (1962) “Wittgenstein on explanation and description”, *The Journal of Philosophy* 59.19, 523 – 530.

Gunnarsson, Logi (2000) *Wittgensteins Leiter: Betrachtungen zum Tractatus*, Berlin: Philo.

Kluge, Friedrich (2012) *Etymologisches Wörterbuch der deutschen Sprache*, Berlin: de Gruyter.

Tatievskaya, Elena (2013) *Wittgenstein über das Verstehen*, Berlin: de Gruyter.

Venieri, Maria (1989) *Wittgenstein über philosophische Erklärung*, Frankfurt a. M.: Peter Lang.

Winch, Peter (1969) “Introduction: the Unity of Wittgenstein’s Philosophy”, *Studies in the Philosophy of Wittgenstein*, 1–19.

Werkausgabe Band 1. Tractatus logico-philosophicus / Tagebücher 1914–1916 / Philosophische Untersuchungen. Frankfurt am Main: Suhrkamp 1984.

Werkausgabe Band 5. Das Blaue Buch / Eine Philosophische Betrachtung (Das Braune Buch). Frankfurt am Main: Suhrkamp 1984.

Wittgenstein and Schopenhauer on Self and Will

Duncan Richter (Lexington, USA)

Abstract

Christopher Janaway argues in a recent essay that Schopenhauer is best understood as using the notion of *will* in a variety of ways. Relatedly, Janaway identifies at least three different ways in which Schopenhauer asks us to think of the self. What Janaway offers is a controversial reconstruction of Schopenhauer's view, but one that I find plausible. I argue in this paper that Wittgenstein offers an interestingly similar, though by no means identical, view of the will in the *Tractatus*. Having set out what Schopenhauer says and what Wittgenstein says, I compare their views and sketch some possible conclusions about the significance of, especially, what Wittgenstein appears to think. These conclusions relate to questions about scientism, metaphysics, the importance of language in philosophy, and the fact/value distinction. That is all too much for a paper of this length to do more than speculate about, but I do hope to show that two of Schopenhauer's notions of self are similar to two of Wittgenstein's notions of will, and that the differences between the two philosophers on this issue raise interesting questions.

In a recent paper on “Different Kinds of Willing in Schopenhauer” (Janaway 2022), Christopher Janaway aims to solve a number of problems about what Schopenhauer says about the negation of the will. To my mind he is successful, but I cannot defend my evaluation of his argument in this paper. Instead, I will explain what Janaway's reconstructed version of Schopenhauer says about the will and the self, and then compare this with what Wittgenstein says about these things in the *Tractatus*. In order to make the best sense of the various claims that Schopenhauer makes about the will, Janaway reconstructs his argument in a doubtless controversial way. This involves identifying three different senses of will. Wittgenstein also talks about three different wills or subjects in the *Tractatus* (the notions of will, self and subject are closely related, as we shall see). I will argue that there is some overlap between Wittgenstein and Janaway's version of Schopenhauer. They are not identical, but comparing the two can be instructive as a way to see more clearly what Wittgenstein is doing in the *Tractatus*. Making the comparison also seems worthwhile given Wittgenstein's early interest in Schopenhauer's philosophy. Any similarities between their ideas are probably not mere coincidences, and differences could be implicit criticisms of Schopenhauer.

A short summary of Schopenhauer's view is in order here. As he sees it, the world can be correctly thought of as both representation or ideas and as will, a

kind of blind striving. Each of us is will, specifically the will to live, and is doomed to suffering and boredom. We suffer when we lack what we are striving for, and when we get it we soon become bored. If we can overcome the will, however, then we can escape this fate. Sadly, how to achieve this desirable goal, and what it could even mean to overcome the will, are not clear. This is where Janaway's reconstruction comes in.

Janaway is especially concerned with questions about how the will can deny itself. If I *am* in some sense the will to live, then how can I ever become will-less? Yet Schopenhauer apparently thinks both that each of us is will and that we should be aiming for complete negation or denial of the will. Since he does not mean that we ought to commit suicide it is unclear what he could possibly mean. Janaway's radical solution is to ignore or re-interpret passages in which Schopenhauer advocates complete negation of the will and instead to see Schopenhauer as calling only for the end of one kind of willing. This, Janaway claims, has three advantages: it fits better with other passages in Schopenhauer's work, it clarifies the connection between moral goodness and liberation (*Erlösung*), and it makes Schopenhauer's view closer to Buddhism which, according to Schopenhauer himself, it is meant to resemble (see Janaway 2022: 11).

Schopenhauer variously characterizes liberation as requiring denial of the self, the overturning of human nature, standing in opposition to nature, and negating the will. Apparently, these ways to attain liberation are all supposed to mean the same thing, but it is not clear that the terms used in them are really synonymous. Janaway makes the "heretical" claim that Schopenhauer's reference to "negation of the will" is "the *least* satisfactory" of these formulations "because *will* is an elusive and unfocused concept, made unsatisfactory by the fact that Schopenhauer puts it to an excessive number of uses" (Janaway 2022: 14).

Janaway argues that Schopenhauer "asks us to think of the self in at least three ways." (Janaway 2022: 15) The first of these is "a living psychophysical individual of the human species" (Janaway 2022: 15). The second is "the real, nonindividuated thing in itself" (Janaway 2022: 15). And the third is "the self as *subject*" (Janaway 2022: 16). This is a difficult notion, though, because "*Subject* is not an ontological category" (Janaway 2022: 16).

In the negation of the will or self, Janaway argues, it cannot be the self as subject (self in the third sense identified above) that is denied, because the liberated self is still aware, and so must still be a subject. Nor can the essential self (self in the second sense) be denied because it is eternal, noumenal. So the self that we are to overcome, negate or deny can only be the particular will of an individual human being (self in the first sense). I do not cease to exist if I achieve liberation, but I do overcome and forego my ego. My dispositions thus change. I no longer care about my own happiness or well-being.

It does not follow, however, that I no longer will at all. I might, for instance, will the well-being of others. This is what Schopenhauer calls compassion in *On the Basis of Morals* (Schopenhauer 2009: 201, see Janaway 2022: 21). Thus, if we follow Janaway and reject Schopenhauer's claim that the will is simply the same thing as the will to life, as well as his claim that liberation needs all willing to cease, then we can reconstruct a version of Schopenhauer that avoids some of the biggest problems his theory seems to run into while being true to views he espouses, albeit he perhaps puts these views forward less often or less prominently than the ones we need to jettison in order to accept Janaway's Schopenhauer. There is something Wittgensteinian about the motivation behind Janaway's reconstruction, too, since Wittgenstein said, according to Elizabeth Anscombe and Peter Geach, that he could not make sense of the idea of the world as will (see Anscombe 1963: 11 and Geach 1957: 558). Attempts to clean up Schopenhauer's notion of will might help us make better sense of his position than Wittgenstein was able to do.

There is some similarity between the three selves of Janaway's Schopenhauer and the three senses of will that Wittgenstein writes about in the *Tractatus*. I will try now to explain what these are. In TLP 5.631 he says that

There is no thinking, representing subject.

If I wrote a book *The World as I Found It* then I would also have to report on my body in it and say which parts are subject to my will and which not, etc., that is to say, this is a means to isolate the subject, or rather to show that in an important sense there is no subject: Of it alone, that is, could there *not* be talk in this book. –

This appears to be at least very close to the idea we find in Schopenhauer of the self as subject. As noted above, subject is not an ontological category, so it makes sense to say, as Wittgenstein does here, that “in an important sense there is no subject.” Nevertheless, we can still talk about it, as long as we are careful about what we are saying. This self, or self in this sense, is not the body or the phenomenal will. Instead, it is that for which the world might be thought to exist as representation. It is not part of the world as anyone finds it, but it is what might be thought to do such finding. It is what Wittgenstein calls the “philosophical I” and the “metaphysical subject” in 5.641.

This metaphysical subject has been identified by Jordi Fairhurst (Fairhurst 2019) and Oskari Kuusela (Kuusela 2017: 45) as the ethical will (what Wittgenstein calls “the will as the bearer of the ethical” in 6.423). 6.423 tells us that nothing can be said of this will, but contrasts it with the phenomenal will, which is of interest only to psychology. Although nothing can be said of the ethical will, Wittgenstein also says that it changes nothing that can be expressed through language. If it changes anything, it changes “the limits of the world, not the facts” (6.43).

We can thus find three wills, or senses of ‘will’, in the *Tractatus*: the metaphysical, the ethical, and the phenomenal. When Schopenhauer talks about negating the will it seems to be some feature of the phenomenal will that he has in mind. Specifically, he opposes the egoism of this will. What, though, are we to make of the metaphysical and ethical will? Wittgenstein’s metaphysical will might be the same thing as the ethical will, as Kuusela and Fairhurst argue, but so far I have done nothing to show that this is the case. Nor have I explained what it would mean.

The metaphysical subject of 5.631 is not only a subject but a subject with a will. What it is is isolated, brought clearly into view, by a process of not only listing the contents of the world but also identifying which of them are subject to this will and which are not. This subject not only perceives but also has attitudes. It wants, trying to move this or that body part, say, or to interact with the world in some other way. This requires choices, which in turn involve evaluation. Some features of the world will be more relevant to it than others. Some will pose threats to the securing of its preferences while others will appear to be beneficial.

These preferences need not be egoistic, but it does seem as though they must exist in some form. Genuinely perceiving the world, we may well think, is not the same as being an indifferent witness to a parade of stimuli. Distinguishing one object or event from another seems to involve some notion of relevant differences, which seems to require a sense of significance or importance. Even if we reject this idea and think that a perceiving subject could be quite indifferent as to what it perceives, *agency* surely involves will (by definition of what constitutes an action) and will involves goals and values of some kind.

Wittgenstein implies that intelligibility involves agency, since language is something that we create (see 2.1, for instance) and use, and that has no meaning otherwise. And a language with no meaning is surely no language at all. So it makes sense for the metaphysical subject, the perceiver of the world, not to be wholly distinct from the ethical subject, the evaluator of objects and events, including its own actions, in the world. Indeed, the implication of the view I have sketched here is that perception and evaluation are not really separable, even if we can sometimes make sense of the idea of merely perceiving or understanding without any judgment or evaluation. If this is correct, and if it is also the case that the subject does not really exist as a thing, as one of the things one might find in the world, then it makes all the more sense to regard the metaphysical subject and the ethical subject simply as one and the same. So far, that is, as it makes sense to count either of them as even one thing, since “in an important sense there is no subject” (TLP 5.631). In other words, although neither the metaphysical perceiver of the world nor the ethical evaluator of the world is an object that might be perceived or evaluated, if we are to talk about one we seem implicitly to be talking about the other as well. It is hard, if not impossible, to conceive of a perceiver that is not an evaluator, and *vice versa*.

This suggests multiple possible conclusions. One is that metaphysics and ethics, or description and evaluation, are not really distinct. Another is that attempts to treat them as distinct, to try to describe the world in a neutral way, as science arguably does, for instance, are to some extent, at least, to be resisted. Non-evaluative description might be possible within certain limits, as acknowledged above, but not universally. It certainly should not be an ideal to which we always aspire. Equally, the fact/value distinction, which the *Tractatus* often seems to defend, would actually be a mistake or, at best,

something to be used with caution and, again, only within certain limits. Finally, since intelligibility or language itself seems to require both evaluation and the identification of distinct objects or events, and since the metaphysical and ethical subjects in an important sense do not exist, philosophers should move away from Schopenhauerian metaphysics and ethics toward considerations of language instead. This certainly sounds like what Wittgenstein does in both his early and his later work.

I can only suggest these conclusions here and not argue properly for them. However, it is at least superficially plausible that Wittgenstein, even early on, would reject both scientism and metaphysics. Further support for some of the claims made here can be found also in work by Anne-Marie Søndergaard Christensen (Christensen 2011: 804-805), Eli Friedlander (Friedlander 2017), and Modesto Gómez-Alonso (Gómez-Alonso 2016).

Let me try to recap what I have said as far as comparing and contrasting Schopenhauer and Wittgenstein on the self or will goes. Schopenhauer, at least as Janaway suggests we read him, distinguishes between the observable human being, the noumenal self, and the self as subject of experience. In the *Tractatus* Wittgenstein also distinguishes between the phenomenal self, the metaphysical self, and the ethical self. The first in each case (the familiar, observable being) is a similar, albeit perhaps not very interesting, idea. The second might sound similar, but in fact Wittgenstein has no use for the notion of a noumenal self. Instead, his metaphysical subject (or will or self) appears to be the same non-thing as his ethical subject (or will or self). This non-thing is equivalent to Schopenhauer's self as subject. So two of the three are at least close to being equivalent. Where Schopenhauer gets blatantly metaphysical and rather mysterious, however, Wittgenstein rejects the quest for discoveries about the nature of reality in philosophy (see TLP 4.111-4.112, for instance) in favour of the activity of clarifying thoughts. This is done from within thought, from within language. Hence if the very meaning of 'action' involves will then where there is action, such as thinking or using language, there is will. This is not metaphysics, though, but what Wittgenstein would later call grammar. It is not a matter of discovery but of bringing more clearly into the light something that we were already, however dimly, aware of.

References

- Anscombe, G. E. M. (1963) *An Introduction to Wittgenstein's Tractatus*, Second Edition, Revised, New York: Harper & Row.
- Christensen, Anne-Marie Søndergaard (2011) "Wittgenstein and Ethics" In Oskari Kuusela and Marie McGinn (eds.) *The Oxford Handbook of Wittgenstein*, Oxford: Oxford University Press, 796-817.
- Fairhurst, Jordi (2019) "The Ethical Subject and the Willing Subject in the Tractatus: an Alternative to the Transcendental Reading", *Philosophia* 47 (1), 75-95.
- Friedlander, Eli (2017) "Logic, Ethics and Existence in Wittgenstein's *Tractatus*", in: Reshef Agam-Segal and Edmund Dain (eds) *Wittgenstein's Moral Thought*, New York: Routledge, 97-131.
- Geach, P. T. (1957) Review of *Ludwig Wittgenstein: Tractatus Logico-Philosophicus* by G. C. M. Colombo, S. J., *The Philosophical Review* 66 (4), 556-559.
- Gómez-Alonso, Modesto (2016) "Wittgenstein on the Will and Voluntary Action", in: Jesús Padilla Gálvez (ed.) *Action, Decision-Making and Forms of Life*, Berlin, Boston: De Gruyter, 77-108.
- Janaway, Christopher (2022) "Different Kinds of Willing in Schopenhauer", in: Judith Norman and Alistair Welchman (eds) *Schopenhauer's 'The World as Will and Representation': A Critical Guide*, Cambridge: Cambridge University Press, 11-25.
- Kuusela, Oskari (2017) "Wittgenstein, Ethics and Philosophical Clarification", in: Edmund Dain and Reshef Agam-Segal (eds) *Wittgenstein's Moral Thought*, New York: Routledge, 37-65.
- Richter, Duncan (2021) *Wittgenstein's Tractatus: A Student's Edition*, Lanham, Maryland: Lexington Books.
- Schopenhauer, Arthur (2009) *The Two Fundamental Problems of Ethics*, translated and edited by Christopher Janaway, Cambridge: Cambridge University Press.

Metaphysics qua Metaphysics

Amedeo Robiolio (London, UK)

Abstract

Metaphysics has never returned to be part of analytic philosophy in the way it is generally said it has, not even in the main neo-Aristotelian proposals. For Aristotle, metaphysics is the study of being qua being. Logical empiricists simply rejected that this is a meaningful study — all that philosophy could do was study representations, be they linguistic or conceptual. Quine is often thought to have refuted this, but I argue that Price has shown that post-Quine metaphysics, e.g. Lewis's, is still a study of representations. However, neo-Aristotelians, e.g. Fine, may seem immune from Price's objection. But I argue that they are not, because they still maintain that only transparent concepts can properly belong to metaphysics. This shows that their metaphysics still studies being in as far as it is represented, not in as far as it is. If metaphysics is possible at all — which is not clear — it would need to allow the study of being through opaque concepts too. Whether plausible or not, the approach that comes closest is Aquinas' notion of 'secundum rem' knowledge.

0. Introduction

The intention of this paper is to pose a challenge: in what sense has analytic philosophy been doing metaphysics? The method that would seem best suited to answer this question is to look at all the plausible available definitions of metaphysics, and then see whether the current analytic tradition can be said to be capable of offering contributions to metaphysics in that sense. For this paper we will be limited to just one sense of the notion of metaphysics, the Aristotelian definition of metaphysics as the study of 'being qua being' (Met. IV, 1003a). But we will not dwell on Aristotelian exegesis, or even on the question of whether Aristotle's own metaphysics respects this definition. All that is needed is that being is studied under only the aspect that it 'is'. This paper concludes that none of the main attempts in analytic philosophy which are generally considered to be metaphysics meet the Aristotelian description, primarily because, as Aristotle contrasts with 'true metaphysics', 'they divide off some portion of it and study the attribute of this portion, as do for example the mathematical sciences', and specifically, the portion they divide off and study is the conceptual or linguistic representation of being, even when this is purportedly abandoned and rejected. I will introduce the arguments showing why logical empiricists rejected metaphysics and why Quine's response is ineffective, and then show, with a Thomistic-inspired objection, how even the neo-Aristotelian approach falls short.

1. Scientific reductions are representational and not ontological

It is well known that the origins of analytic philosophy were characterised by a markedly anti-metaphysical sentiment. The usual way this is introduced is by appeal to some form of positivism, empiricism, verificationism and logicism. What these have in common is that they establish a certain criterion to determine what can be true — i.e. ‘truth-functionality’. This criterion is best captured by the statement that a proposition is truth functional if and only if it can be expressed through a scientific reduction.

Van Gulik (2001) proposes that reduction is captured by the “nothing more than ...” slogan and provides a crucial distinction, between ‘ontological’ and ‘representational’ reduction. To him, the ‘logical empiricist’ position, which captures early analytic antimetaphysicians, falls within the supporters of representational reduction (p. 10). This seems right. For the ontological reductionist, it is entirely possible that while science tells us that *x* is nothing more than *y*, in reality — which might be beyond our reach — *x* is nothing more than *z*. For the representational reductionist, however, this is a confusion of terms. If science tells us that *x* is nothing more than *y*, then this is the very meaning of the term *x*, or we might say alternatively, the content of the concept. We might also say that there cannot be a notion of truth that goes beyond what is made true by our reasoning and observation. This seems to be what early analytic philosophy had in mind.

It is then no coincidence that the anti-metaphysical phase of analytic philosophy is often described in terms of the ‘linguistic turn’. It is possible to generalise a little and substitute ‘representation’ for ‘language’, such as to also include the more conceptualistic aspects of the linguistic turn. At core is a collapse of reality into representation, since what makes our sentences or conceptions true is their relation to our theories and models, not their relation to some supposed reality beyond them. Then, a study of being qua being is senseless. The question is then when, if ever at all, metaphysics in the ‘being qua being’ sense started in the analytic sphere.

2. If to be is to be the value of a variable, then being is studied only representationally

That analytic philosophy has abandoned its consciously anti-metaphysical sentiment is quite evident. Whether it has abandoned its focus on the study of being qua its representation and its avoidance of the study of being qua being, however, is an entirely different question. There might have been a mere change of label, with a much less significant change of content. This is the conclusion which Price (2009) demonstrated. To understand his argument, it should first be made clear what brought metaphysics back into analytic philosophy. The typical answer is that it was Quine, who, making ontology analysable with the slogan: ‘To be is, purely and simply, to be the value of a variable’(1948 p. 32) and undermining the notion of scientific reduction that had banished metaphysics (1951), ushered in the revival of metaphysics.

Price’s first argument, challenging the import of Two Dogmas, rests on the premiss that “the traditional metaphysician wants to be able to say, ‘I agree it is useful to say this, but is it true?’” (p. 6). Price wants to show that this premiss is never fulfilled, and Quine does not achieve this for the traditional metaphysician. Quine never really offers anything that goes beyond the pragmatic, he merely rejects that there is a clear distinction between the empirically and the logically pragmatic. Price therefore maintains that ‘Quine is not returning to the kind of metaphysics rejected by the logical empiricists. On the contrary, he is moving forwards, embracing a more thoroughgoing post-positivist pragmatism.’ (p. 7).

The relation of this to the issue of being and representation is not touched upon directly by Price. But the notion of pragmatism Price introduces encapsulates the way in which Quine has not moved at all from metaphysics as the study of being qua its representation. The main thesis of “On what there is” proves this. To link being to the value of a variable, i.e. an object of quantification, is analogous to the logical empiricists taking the scientific and linguistic (or conceptual) model of the world to in fact be the world. For the value of a variable is merely an item of a representation, and not what the traditional metaphysician would intuitively think of. There is no appeal to being as something representation-independent, thus ontological commitment is concerned with the proper rules of “is” and “exists” phrases, not with being qua being.

The account given here by Price needs disambiguation. Early analytic philosophy had provided a critique of metaphysics, and Quine's rebuttal did little to restore it, because, essentially, he did not overcome the purely representational way it relates to being. But here we could say that current metaphysics ignores the shortcoming of Quine's argument and happily talks about being qua being: in this sense metaphysics is a ghost who walks because it does more than Quine's arguments allow. Or, that current metaphysics is doing exactly as little as Quine's argument allowed, and is under the impression that that is the study of being qua being: then metaphysics is a ghost who walks because it thinks, mistakenly, that what Quine allows is the study of being qua being.

The latter seems more plausible. To show this, an example might be found by attending to Lewis 1986. Here Price is clear about this conclusion: "Many people think [*italics original*] that they are doing metaphysics, in Lewis's footsteps ..." (Price 2009 pp. 20-21). The point is: Lewis has always advocated for modal realism not out of the conviction that possible worlds are real, but out of the conviction that a system of reasoning where possible worlds are real works much better (Lewis 1986, vii) than one where they are not. It is a pragmatic issue, in essentially the same way as the logical empiricists thought of truth in scientific models. The attention to the issue of representation which has been advocated for here makes this fit even tighter: Lewis is interested in providing us with a good linguistic and conceptual representation, and not with a study of being qua being.

3. If metaphysical concepts are transparent, metaphysics only studies representations

The Quinean and Lewisian approach to ontology has been challenged also by those on the other side, thinking that while perhaps it did study being qua being, it only did so in a way that did not allow for structural richness, e.g. not allowing to distinguish essence and modality, and not containing a hierarchy of ontological categories. Those proposals are labelled neo-Aristotelians. The question is whether these neo-Aristotelian philosophers have finally thrown off the habit of treating metaphysics as the study of being qua its representation to engage with being qua being.

The argument for the negative answer is grounded in the fact that neo-Aristotelian analytic philosophers believe metaphysical concepts to be ‘transparent’. The idea is that this transparency is yet another iteration of the reality-representation collapse proposed by the logical empiricist analytical philosophy, in which once again the representation is the one winning over. The argument for the transparency of metaphysical concepts is made clearest by Kit Fine (2011). In this chapter, as well as in much more of the neoAristotelian literature, there is a clear intention to study being qua being — thus for example “generality” is one of his desiderata (p. 16). But because of transparency, this seems insufficient.

Fine’s argument for transparency aims to establish this: “a concept is transparent if there is no significant gap between the concept and what it is a concept of.” (Fine 2011 p. 9) That is, in contrast with opaque concept-object relations, such as the concept-water and the object H₂O relation, there are transparent concept-object relations. Which, it appears unlike water and H₂O, are the proper subject matter of metaphysics: and that is because there is ‘no significant gap between the concept identity and the identity relation of which it is a concept’ (ibid).

His argument to this effect is grounded in the assumption that metaphysics is made up of a priori necessary truths. He distinguishes the transparency of metaphysical concepts into modal and epistemic transparency. Epistemic transparency is the notion that ‘every necessary truth is a priori’ (id. p. 22). Modal transparency instead claims that ‘some concepts are necessarily the concept of their object’ — and only those, it seems, can be metaphysical concepts (id. p. 23). ‘Water is H₂O’ is necessary, but the concept ‘water’ is not necessarily the concept of H₂O, so Fine regards that as modally opaque. Since ‘water is H₂O’ is considered necessary but a posteriori, it is really only within the realm of modally transparent concepts that the universality of epistemic transparency holds.

What is called ‘concept’ is what has been presented as representation so far. To find out if the study of being qua being just is the study of a priori truths about modally transparent concepts, we must wonder whether there might not be a posteriori metaphysical truths. To deny that this question makes sense would be analogous to the logical empiricist’s rejection of the question of whether

there might be a gap between reality and the truths built in the language of a scientific reduction.

The very example Fine chose, i.e. that there is ‘no significant gap between the concept identity and the identity relation of which it is a concept.’ (p. 9) had been criticised extensively, precisely under the influence of Aristotle. This criticism is part of Aquinas’ *secundum rem* (according to the thing) / *secundum rationem* (according to reason) distinction. Aquinas distinguishes *secundum rem* and *secundum rationem* for many themes, and the broader worry is that by postulating the transparency of metaphysical concepts we remain confined to the *secundum rationem* realm, and completely oblivious to the *secundum rem* realm, even though studying it may be the study of being qua being.

Hughes (1989) summarises Aquinas’ position: ‘Aquinas maintains that although a thing is the same as itself realiter, and not just *secundum rationem*, the relation of self-identity exists only *secundum rationem*’ (p. 220). It is tempting to associate the *secundum rationem* with the ‘qua being’ study of being. Hence, we should infer that the concept of identity which we access a priori is merely the *secundum rationem* one. For it seems clear that the study of being qua being should be constituted of *secundum rem* truths, and the study of being qua its representation, as *secundum rationem* truths.

The sceptic might ask how we know anything *secundum rem*. But this is no different than asking how we know that ‘water = H₂O’ is necessarily true. The most likely answer is Aquinas’ conception of knowledge as ‘adequation between the intellect and a thing’ (Wippel 1989, p. 295). We might say that truth is grounded both in the concept and in the being of things, but that Aquinas thinks being takes primacy. A priori knowledge is however exclusively grounded in the concept. However, by adequation we enter into contact also with the existence of necessary truths, when they are instantiated — in the case of self-identity, anything is an example. This would mean that to study being qua being we must face modally opaque truths.

The neo-Aristotelians do not have this in mind. Stuck in the a priori grounding of truths, they are cut out from studying being qua being. This might be because Aquinas’s epistemology is implausible: but this would only mean that one cannot study ‘being qua being’.

This does not prove the possibility of a study of being qua being. But it is advocating for clarity when it comes to the use of the word ‘metaphysics’ in analytic philosophy: so far it seems that under that label, when it was used at all, it provided only the study of being qua its representation, but on at least two occasions, exemplified here by Lewis and Fine, this is not made clear. If anyone wanted an investigation of being qua being, current analytic metaphysics will not deliver it. If this is because it has proven impossible, it is important that this is clear. If this is because analytic philosophy thought it delivered it when it did not, better attempts are necessary.

References

- Fine, Kit (2011) "What is metaphysics?", in: Tuomas E. Tahko (ed.), *Contemporary Aristotelian Metaphysics*, Cambridge University Press, 8–25.
- Hughes, Christopher (1989) *On a Complex Theory of a Simple God: An Investigation in Aquinas' Philosophical Theology*, Cornell University Press.
- Lewis, David (1986) *On the Plurality of Worlds*, Wiley-Blackwell.
- Price, Huw (2009) "Metaphysics after Carnap : the ghost who walks?", in: David Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford University Press, 320–46.
- Quine, W. V. O. (1948) "On what there is", *The Review of Metaphysics*, Vol. 2, No. 5, 21-38.
- Quine, W. V. O. (1951). "Two Dogmas of Empiricism", *Philosophical Review* 60 (1), 20–43.
- Van Gulick, Robert (2001) "Reduction, emergence and other recent options on the mind/body problem: A philosophic overview", *Journal of Consciousness Studies* 8 (9-10), 1-34.
- Wippel, John F. (1989) "Truth in Thomas Aquinas", *The Review of Metaphysics*, vol. 43, no. 2, 295–326.

A Puzzle About Grounding and Specification

Pauliina Rumm (Cambridge, USA)

Abstract

It is common to think that the species-genus relation and the determinate-determinable relation are in some tight way connected to metaphysical grounding. Gideon Rosen (2010) has proposed two principles to capture those connections that have become widely accepted. In the case of the determinate-determinable relation, he argues that facts about determinables are grounded in facts about determinates. For example, the fact that my shirt is red is grounded in its being crimson, which is a determinate tone of red. In the case of the species-genus relation, however, he argues the grounding relation runs from the more general to the more specific: facts about species are grounded in facts about their genus and differentia. For example, the fact that something is a square is grounded in its being a rectangle and in its being equilateral. But while such principles are extremely plausible, I argue that they give rise to viciously circular grounding structures. I propose that the problem arises due to an unjustifiably tight relation Rosen takes there to be between grounding and the determinate-determinable relation. For as I will show, it is plausible to think that, differently from the species-genus relation, the determinate-determinable relation can be captured in essentialist terms, without any mention of grounding. Consequently, I think the differences between the species-genus and the determinate-determinable relation cannot be understood in terms of grounding.

There are two particularly well-known specification relations in metaphysics: the determinate-determinable relation and the species-genus relation. The two are importantly different. If F is a determinate of G , then being F is a more specific way to be G . For example, the property *being cerulean* is determinate of the property *being blue* in part because being cerulean is a more specific way to be blue. Importantly, a determinate must always characterize objects in respect of its determinables. For example, *being cerulean* characterizes objects in respect of their color, which distinguishes it from properties, such as *being colored and square*, that are perhaps in some sense more specific than *being colored* but are not determinates of it (Funkhouser 2006, Prior 1949). In contrast, the species-genus relation is concerned with how certain entities can be defined in terms of more general ones. On the traditional way of understanding the relation, “a species is defined as the conjunction of genus and differentia” (Rosen 2010: 127). For example, *square* is a species of *rectangle* because a square just is an equilateral (differentia) rectangle (genus). Here, 'definition' should be understood as a real definition, not a nominal one (Rosen 2015).

1. The Puzzle

As with many metaphysical relations, it is intuitive to think that both the determinate-determinable and the species-genus relation are closely related to metaphysical grounding, which is a non-causal explanatory relation that is commonly picked out by words, such as 'grounds', 'makes it the case that', and 'because'. In particular, it is natural to accept two principles, originally proposed by Gideon Rosen (2010): the Determinate-Determinable Link and the Species-Genus Link. According to the Determinate-Determinable Link, the fact that some object instantiates a determinable is always grounded in its instantiating some determinate property. This seems very plausible. It simply feels natural to say that the fact that my shirt is blue is grounded in its being cerulean or that my being more than 100 cm tall is grounded in my being 178 cm tall. The Species-Genus Link is similarly plausible. According to it, the fact that some object instantiates a species is grounded in its instantiating a differentia and a genus. For example, the fact that something is a square is grounded in its being equilateral and in its being a rectangle. Again, such a grounding claim simply strikes us as plausible.

But while both the Determinate-Determinable Link and the Species-Genus Link seem intuitive and have come to be widely accepted (e.g., Ford 2021, Schaffer 2012, Shumener 2022), they seem to give rise to viciously circular grounding structures. To see why, consider the following claims:

1. The relation *being a child of* is a determinate of the determinable *being related to*.
2. The relation *being a daughter of* is a determinate of the determinable *being related to*.
3. The relation *being a daughter of* is a species of the genus *being a child of* with *being female* as the differentia.
4. If x and y are both determinates of a determinable z , and x entails y (but not vice versa), then x is a determinate of y .

While all the claims seem plausible, they cannot be jointly true if we are right about how the determinate-determinable relation and the species-genus relation are related to grounding. For then we would have to hold the following grounding claims, which give rise to a circle of ground:

- [Jill is Jack's daughter] \leftarrow [Jill is Jack's child], [Jill is female]
- [Jill is Jack's child] \leftarrow [Jill is Jack's daughter]

where '[*p*]' stands for 'the fact that *p*' and ' \leftarrow ' stands for 'is grounded in'. The first claim follows from (3) and the Species-Genus Link. The second claim follows from (1), (2), (4) and the Determinate-Determinable Link. But together, the two claims commit us to holding that Jill is Jack's daughter in part because she is Jack's child and she is Jack's child because she is his daughter. That seems implausible. For even if one denies that grounding in general is asymmetric, it is not plausible to think that this is one of the cases where asymmetry fails.

It is also worth noting that the problem is not specific to kinship relations, but is more general. For example, we can generate analogously circular grounding structures by thinking about certain shape properties. To see how, consider the following claims:

1. The property *being a square* is a determinate of the determinable *being a shape*.
2. The property *being a rectangle* is a determinate of the determinable *being a shape*.
3. The property *being a square* is a species of the genus *being a rectangle* with *being equilateral* as the differentia.
4. If *x* and *y* are both determinates of a determinable *z*, and *x* entails *y* (but not vice versa), then *x* is a determinate of *y*.

Like before, all four principles seem plausible but, if the Determinate-Determinable Link and the Species-Genus Link are correct, we again have to accept the following two claims that give rise to a circular grounding structure:

- [*ABCD* is a square] \leftarrow [*ABCD* is a rectangle], [*ABCD* is equilateral]
- [*ABCD* is a rectangle] \leftarrow [*ABCD* is a square]

Consequently, since such circular grounding structures seem to arise more generally and are not specific to specific kinds of properties, they suggest that something has gone wrong in how we think about the relation between the different kinds of specification relations and grounding.

At first glance, it might seem most plausible to simply reject (4), according to which, if x and y are both determinates of a determinable z , and x entails y (but not vice versa), then x is a determinate of y . However, rejecting the principle due to such cases seems *ad hoc* since the principle seems in general extremely plausible. For example, simply from knowing that crimson and red are both determinates of red and that being crimson entails being red, it seems we can correctly infer that crimson must itself be a determinate of red. Or similarly, just from knowing that weighing 50 kgs and weighing more than 5 kgs are both determinates of *having weight* and that weighing 50 kgs entails weighing more than 5 kgs, we can correctly infer that weighing 50 kgs is a determinate of weighing more than 5 kgs. It would thus seem unmotivated to claim that, although (4) holds in almost all of the canonical cases of the determinate-determinable relation, it nevertheless just so happens to fail whenever there is a case that threatens to give rise to a circle of ground; and so, we should try to avoid giving up (4) for as long as possible.

It would be similarly implausible to try to solve the puzzle by rejecting claims (1)–(3) in either of the cases I considered above. For one thing, again, the claims are simply very intuitive. For instance, it seems just natural to think of *being a child of* and of *being a son of* as just specifications of the relation *being related to*. But for another thing, as I have already suggested, the puzzle does not seem to be specific to only specific kinds of relations; and so, it would be unattractively piecemeal to try to avoid the puzzle by arguing why, in each of the problem cases, there is some distinct reason that undermines our seemingly correct judgments about them.

But this leaves us with two options: either we must reject the Determinate-Determinable Link or the Species-Genus Link. It seems to me that rejecting the latter is implausible since many hold that real definition should be understood in terms of grounding, in which case the Species-Genus Link follows directly from what it takes for the species-genus relation to hold in the first place (Rosen 2015). Instead, I think we should reject the Determinate-Determinable

Link. For as I will now show, differently from the species-genus relation, it is plausible to think that a proper analysis of the determinate-determinable relation does not turn on ground-theoretic notions and thus gives us no reason to think it is related to grounding in the way the Determinate-Determinable Link takes it to be.

2. Giving Up the Determinate-Determinable Link

There are many well-known accounts of the determinate-determinable relation that do not make reference to grounding and could thus be used to motivate giving up the Determinate-Determinable Link. However, since an appropriate solution to the puzzle should also explain why the Determinate-Determinable Link seemed plausible in the first place, most of the proposed accounts, such as those relying on facts about property spaces or causal powers, would not do. A different account is needed that relies on a notion that we naturally take to be tightly related to grounding relations. Let me provide such an account here.

As I mentioned earlier, a core part of F 's being a determinate of G is that an object's being F is a more specific way for it to be G ; and moreover, an object's being F characterizes the object in respect of G . What that amounts to is, I think, best captured by the following analysis. F is a determinate of G iff:

1. For every proposition p , if it is essential to an arbitrary object's being G that p is the case, it is also essential to its being F that p is the case.
2. For every proposition p , if it is essential to an arbitrary object's being identical to G that p is the case, it is also essential to its being identical to F that p is the case, but not vice versa.

While these conditions are closely related to Stephen Yablo's (1992) essentialist analysis of the determinate-determinable relation, I will not rely on the modal conception of essence Yablo relies on. Instead, I will understand essence as a primitive notion that characterizes what it is, *at its very core*, to have a certain property or to be identical to a certain entity (Fine 2015). For example, it is essential to something's being a square that it is a rectangle and that it is equilateral, not because it is necessary, but because that characterizes what it is, at its very core, to be a square. We can make similar claims about *being a*

square, the property itself. For example, it is essential to something's being identical to *being a square* that it is an abstract object. Notably, while all such essential truths are necessary, not all necessary truths are essential. For example, while having the disjunctive property *being a rectangle or being pink* is necessary to an object's being a square, it does not characterize what, at its very core, it is to be a square.

With such a conception of essence in hand, we can see why (1) and (2) provide an attractive analysis of the determinate-determinable relation. First, they hold in the paradigmatic cases of the determinate-determinable relation. For example, consider *cerulean* and *cerulean-25* (a specific tone of cerulean). The first condition is fulfilled since there does not seem to be anything that characterizes what it is for an object to be cerulean but fails to characterize what it is for it to be cerulean-25. The same way it is essential to an arbitrary object's being cerulean that it is blue, that it is colored, etc., it is also essential to its being cerulean-25 that it is blue, that it is colored, etc. Likewise, the second condition is fulfilled since there does not seem anything that is essential to an arbitrary object's being identical to the color *cerulean* but is not essential to its being identical to the color *cerulean-25*. Both of them are essentially color properties, tones of blue, and so on. And yet, there is something that is essential to being identical to *cerulean-25* but not essential to being identical to *cerulean*: namely, being a tone of cerulean. Since similar reasoning applies in other canonical cases of the determinate-determinable relation, the account avoids false negatives.

Relatedly, the analysis avoids certain obvious counterexamples. First, it correctly predicts that it is not always the case that $P \wedge Q$ is a determinate of P or of Q . The reason why, for example, *being red on Thursday* is not a determinate of *being red* is that the second condition fails. For it is essential to something's being identical to the property *being red* that it is a color property, but it is not essential to something's being identical to the property *being red on Thursday* that it is a color property. For if you ask me what my favorite color is and I say, "Red on Thursday," I am obviously confused; *red on Thursday* is not a color. Since the second condition fails similarly in other cases of gerrymandered conjunctive properties, the principle avoids taking any

random conjunctive property to be a determinate of its conjuncts. Indeed, the second condition ensures that a determinate always characterizes an object in respect of its determinables.

Likewise, the account correctly predicts that P and Q are not determinates of $P \vee Q$. For there is always something that is essential to an object's instantiating $P \vee Q$ that is not essential to its instantiating P or to its instantiating Q . To see why, consider the following example. Although it is essential to an object's instantiating *red or round* that it instantiates *red* or that it instantiates *round*, it is not essential to its instantiating *red* that it instantiates *red* or that it instantiates *round*. For what is essential to something's being red should make no reference to roundness, given I can have a complete understanding of redness while knowing nothing about roundness. Thus, the first condition fails to hold in the case of disjunctive properties and their disjuncts.

The essentialist analysis also ensures that other features of the determinate-determinable relation hold. For example, it explains why an object that instantiates some property, F , also instantiates all the determinables of F . The thought behind the essentialist analysis is that, in a certain sense, the essence of each of F 's determinables is contained within the essence of F itself. For example, everything we can say about what, at its very core, it is for an object to be red is already contained within what, at its very core, it is for an object to be crimson. But then it naturally follows that an object cannot be crimson while failing to be red. Since condition (2) introduces an asymmetry between a determinable and its determinates, similar reasoning about conditions (1) and (2) also explains the modal asymmetry of the determinate-determinable relation: if F is a determinate of G , then if x is F then it must also be G but it is possible for some y to be G without being F .

Now, there are some features of the determinate-determinable relation that the above analysis does not account for. For example, it does not explain why same-level determinates, such as *red* and *blue*, exclude one another. But this does not make the account implausible. For not all features that characterize a relation need to be part of or follow from the account of what it takes for the relation to hold. Indeed, the more pressing concern is that the account turns out to be incompatible with the principle that determinates can specify their

determinables only non-conjunctively since it regards many cases of the species-genus relation as cases of the determinate-determinable relation. However, it is not obvious to me why we should consider this implausible. Of course, as my account shows, not *all* cases of conjunctive specification are cases of the determinate-determinable relation, but why think none are? After all, it is commonly thought that, say, *being a square* is a determinate of *being a rectangle* (Funkhouser 2014: 40); and so, the assumption that no instance of the determinate-determinable relation can be an instance of conjunctive specification seems unwarranted.

3. Solving the Puzzle

Returning to the initial puzzle, we can see why giving up the relation between the determinate-determinable relation and grounding is helpful. It allows us to hold that *being a square* is a determinate of *being a rectangle* without thereby having to hold that $[ABCD \text{ is a square}]$ grounds $[ABCD \text{ is a rectangle}]$. The first condition is fulfilled since it seems that there is nothing more to being a rectangle than is already contained in being a square. The same way it is essential to an arbitrary object's being a rectangle that it is a shape, has four sides, etc., it is also essential to its being a square that it is a shape, has four sides, etc. The second condition is similarly fulfilled because everything essential to an object's being identical to the property *being a rectangle* is also essential to its being identical to the property *being a square*. Both are shape properties, abstract objects, and so on. And again, there is something that is essential to being identical to the property *being a square* that is not essential to being identical to *being a rectangle*: namely, being a property that characterizes being a certain kind of rectangle. But since neither of the conditions turns on ground-theoretic notions, we do not need to simultaneously hold that $[ABCD \text{ is a square}]$ grounds $[ABCD \text{ is a rectangle}]$; and so, we get rid of the vicious circle that arose. The same kind of reasoning can also be applied in other instances of the puzzle so that the approach does not seem implausibly piecemeal.

A benefit of solving the puzzle by adopting the essentialist account of the determinate-determinable relation is that it helps to explain why the Determinate-Determinable Link seemed plausible in the first place. The reason

the Determinate-Determinable Link holds in a wide number of cases is due to the intuitive relation that exists between grounding and essence. There certainly seems to be a tight link between what, at its very core, it is to have a certain property and what makes it the case that something has the property. Nevertheless, the connection is not strong enough for the Determinate-Determinable Link to be plausible.

References

Fine, Kit (2015) "Unified Foundations for Essence and Ground", *Journal of the American Philosophical Association* 1 (2), 296-311.

Ford, Anton (2011) "Action and Generality", in: Anton Ford, Jennifer Hornsby, and Frederick Stoutland (eds.) *Essays on Anscombe's Intention*, Cambridge: Harvard University Press, 76–104.

Funkhouser, Eric (2014) *The Logical Structure of Kinds*, Oxford: Oxford University Press.

Funkhouser, Eric (2006) "The Determinable-Determinate Relation", *Noûs* 40 (3), 548–569.

Prior, Arthur N. (1949) "I. Determinables, Determinates and Determinants", *Mind* 58 (229), 1–20.

Rosen, Gideon (2015) "Real Definition", *Analytic Philosophy* 56 (3), 189–209.

Rosen, Gideon (2010) "Metaphysical Dependence: Grounding and Reduction", in: Bob Hale and Aviv Hoffmann (eds.) *Modality: Metaphysics, Logic, and Epistemology*, Oxford: Oxford University Press, 109–135.

Schaffer, Jonathan (2012) "Grounding, Transitivity, and Contrastivity", in: Fabrice Correia and Benjamin Schnieder (eds.) *Metaphysical Grounding: Understanding the Structure of Reality*, Cambridge: Cambridge University Press, 122–138.

Shumener, Erica (2022) "Intrinsicity and Determinacy", *Philosophical Studies* 179 (11), 3349–3364.

Yablo, Stephen (1992) "Mental Causation", *Philosophical Review* 101 (2), 245–280.

Acknowledgement

Many thanks to Selim Berker, Luke Ciancarelli, and Erica Shumener for helpful conversations that inspired many of the arguments in this essay.

Special thanks to Florence Bacus for multiple conversations and written feedback on earlier drafts of the essay and to Selim Berker for pointing out a version of the puzzle.

How to Avoid Metaphysical Magic – From Social Ontology to Mathematics

Jenni Rytilä (Tampere, Finland)

Abstract

Metaphysical views that see the existence of entities, such as mathematical objects or social entities, following easily from certain kinds of truths or facts can face so called magic objections. The worry is that easy or lightweight existence of objects is not legitimate, but only a metaphysical magic trick. This paper tackles magic objections that can be raised against a social constructionist view of mathematics. The idea is that mathematical entities exist due to our mathematical practices, and as abstract social constructions they have a lightweight existence compared to concrete objects and activities. The question is whether positing existing mathematical entities on top of the practices is legitimate or just piece of metaphysical magic. My strategy for responding to magic objections is to use arguments from social ontology and adapt them to the case of socially constructed mathematical entities. We have good reasons to affirm the existence of abstract mathematical entities, but because they are social constructions, their existence does not require anything more mysterious than the appropriate practices. At the same time, the existence conditions for mathematical entities are typically more demanding than for social entities due to the highly constrained nature of mathematical practices. This, together with connections to the natural world, results in entities that are neither arbitrary, nor magic-like.

1. Introduction

It might be said that philosophy of mathematics is in the midst of a social turn, mirroring similar developments in philosophy of science and analytic philosophy. The point of departure for the social turn is that understanding the nature of mathematics, also in philosophical terms, requires looking closely at how we as human beings do mathematics. As a result, the social dimensions of mathematics are increasingly recognized and given significance. One approach where the social turn is taken in full is *social constructionism*. This approach sees mathematics as a fundamentally social phenomenon and uses ideas from social ontology to understand the nature of the practices and the subject matter of mathematics (see e.g., Cole 2015, Feferman 2009, Rytilä 2021).

Social constructionism, understood as a metaphysical view, states that the abstract entities of mathematics (numbers, geometric figures and structures, set-theoretic universe, and so on) are products of our social mathematical practices, but at the same time they are *in some sense* real existing things. They have a similar way of existing as social (and cultural) entities, such as money, corporations, borders, and symphonies. Moreover, mathematical entities so

conceived exist in a "lightweight" way compared to the existence of concrete things or the activities actually performed; the thought is that we get the abstract entities for cheap whenever the appropriate practices exist.

In this paper, I tackle a type of objection that faces such "lightweight" ontologies, where the existence of objects does not require much. The worry is that this kind of existence is not legitimate, but a metaphysical magic trick. Is it not suspicious to posit existing mathematical entities on top of the practices? Are mathematicians just conjuring the objects of their study into being, as if by magic? I will argue that such magic objections can be resisted in the case of socially constructed mathematical entities. My argument expands on a similar case against magic I have previously made in (Hartimo & Rytilä 2023).

I start with presenting the problem by going through some magic objections concerning the existence of both mathematical and social entities. My strategy for responding to magic objections is to use arguments from social ontology and adapt them to mathematics. Two main reasons to deny metaphysical magic in the case of social entities are that their existence does not require much, but at the same time, it is not too easy. Following this line of argumentation, I then give three reasons why socially constructing mathematical entities is not magic: 1) their existence is not too mysterious, 2) existence conditions of mathematical entities are typically more demanding than for social entities, and 3) due to their connections to features of the natural world, mathematical social constructions are not magic-like but real, legitimate entities. To conclude the paper, I examine how my case against magic differs from similar arguments. Whereas others avoid metaphysical magic by denying that mathematical entities are created by our practices and linguistic principles, my point is that this creation is needed but it is not objectionable or mysterious. Thus, I suggest that mathematicians are more like skilful artisans than magicians.

2. Magic objections in mathematics and social ontology

Traditional ontological platonism assumes that mathematical entities exist in a substantial and independent way, analogous to the existence of objects like rocks and stars. In contrast, other views see the existence of mathematical entities following easily from certain kinds of truths. Two representative

examples are Linnebo's (2018) account of thin objects and Rayo's (2015) account of trivial existence. Linnebo argues that the existence of mathematical entities, like directions, numbers, and sets, does not require anything more than the appropriate abstraction principles. This makes them thin objects, because their existence does not make a (further) substantial demand on the world. In a similar vein, Rayo argues that numbers exist trivially, because the truth-conditions of arithmetical statements like '1+1=2' are satisfied trivially, meaning that statement is true regardless of what the world is like.

Thin or trivial existence has the benefit of respecting important features of mathematics – mathematical statements making reference to objects, the abstract nature and infinite number of mathematical entities, mathematical truth being of greater importance than ontology – without incurring problems associated with traditional ontological platonism. However, assuming that numbers and other mathematical entities have a “lightweight” existence, different than the substantial existence of concrete objects, raises a particular worry:

Metaontological minimalism [the view that existence and objecthood do not require much] can come across as a piece of philosophical magic that aspires to conjure up something out of nothing – or, in the relative case, to conjure up more out of less. (Linnebo 2018: 5)

Our feeling of hocus-pocus about the ‘easy’ proof of numbers (etc.) is really very strong and has got to be respected. (Yablo 2000: 199)

We can call these kinds of worries about existence being so easy as to be suspicious magic objections. Magic objections are a worry also for a social constructionist view of mathematical entities. Although social constructionism is in many important ways a different view of mathematics than the accounts of Linnebo and Rayo, it shares the stance on mathematical entities having lightweight existence. The thought is that the existence of mathematical entities does not require substantially more from the world than the existence of the appropriate practices. In other words, we get the abstract entities for cheap on top of the practices. Consequently, the idea of abstract mathematical entities existing as social constructions can face the same objection that we cannot just define things into existence.

Indeed, similar worries have been raised about the existence of more familiar social entities. For example:

I think it is strange that merely speaking and intoning certain phrases could cause anything to exist...The thought is that only wizards and warlocks can bring things into existence by merely uttering a few phrases. (Effingham 2010: 253)

The example here is church parishioners creating a football team by making a verbal agreement, but the worry can also be stated more generally:

Our sense that there is an element of magic, a conjuring trick, a sleight of hand in the creation of institutional facts out of brute facts [...]. (Searle 1995: 45)

The question then is, how to avoid the worry that assuming the existence of mathematical social constructions – in addition to the concrete mathematical practices – is unjustified and suspicious, just a metaphysical magic trick?

3. How to respond to magic objections?

My strategy for responding to magic objections is to use arguments from social ontology and adapt them to mathematics. In the context of social ontology, Thomasson (2019) argues that magic objections are misguided. If there is magic involved in generating social entities, like a climate committee or a corporation, it is such poor magic as to not be magic at all:

Consider the worst magic trick in the world: ‘Nothing up my sleeves... I’ll just put this right glove and this left glove in a hat and... Shazam! A pair of gloves emerges!’ (Thomasson 2019: 4831)

There are two main reasons why the “magic trick” of creating social entities is no magic trick at all. The first is that existence does not require much. According to Thomasson’s easy ontology -approach, all it takes for a certain object to exist is for the term’s application conditions to be fulfilled. The idea is that there is nothing more mysterious required for a corporation to exist than the relevant papers having been filed, just as there is nothing more needed for a pair of gloves to exist than the existence of a left glove and a right glove. To

put the idea another way, the conditions for the existence of a certain social entity is just the appropriate social practice; nothing more could be pointed to if we are asked why corporations or marriages exist. Moreover, denying existence for social entities is often more odd than affirming it, since we successfully deal with and talk about social entities all the time in ordinary life as well as in social sciences. It does not make much sense to do economics and deny the existence of recessions, or at a conference say “I’m not from any university because universities don’t exist”. Accepting the existence of social entities respects our everyday and scientific practices.

The second reason to deny magic in social ontology is that, at the same time, existence is not *too* easy. Usually there is more needed for a social entity to exist than merely saying some relevant “magic words”, and application conditions differ from object to object. Just a declaration may suffice for the existence of a climate committee, but a vote by parliament is needed for a law to exist, and for other things, some underlying physical or historical facts may be needed (Thomasson 2019: 4831).

So, for social entities, we have reasons to affirm their existence and to deny that they are a result of metaphysical magic. I argue that the same reasons can be adapted to mathematics when mathematical entities are seen as social constructions. From this viewpoint, mathematical entities have significant similarities with social entities, and lack some characteristics traditionally attributed to them, such as timeless or necessary existence, that would complicate the case against metaphysical magic.

4. How to avoid metaphysical magic in mathematics?

Following the arguments from social ontology, the first reason why social construction of mathematical objects is no magic trick is that their existence is not too mysterious. We successfully speak about and deal with mathematical entities in our mathematical practices, so they are entitled to some kind of existence. But unlike the substantial existence of concrete things, the lightweight existence of abstract mathematical entities does not require much. There is nothing more needed to bring them into existence than the appropriate practices. Similarly, Pantsar (2024) argues that numbers exist just as referents of culturally shared number concepts. An important benefit of

accepting the (lightweight) existence of mathematical entities is that it allows for face-value readings of mathematical language, meaning there is no need to say we are wrong or do not mean what we say when we make claims about mathematical entities, such as “7 is prime” (see Linnebo 2018).

Moreover, the existence of numbers, functions, triangles, or sets is not more mysterious than the existence of corporations, committees, recessions, and symphonies; they all are socially constructed, existing things. On this point, social constructionism has a benefit over other lightweight views of mathematical existence, such as Linnebo’s thin objects. Linnebo (2018) sees abstraction as productive – new entities come into existence from abstraction on old entities – but at the same time, he holds that mathematical entities are counterfactually independent of us. So there remains a sense that sets and numbers just pop out of nowhere by themselves, magic-like. In contrast, the existence of socially constructed mathematical entities is less mysterious, because they are produced by us and our practices and concepts.

On the other hand, not just anything is brought into existence by our social practices of mathematics. The second reason to avoid magic is that the existence of mathematical entities is not too easy; in fact the application conditions for mathematical entities – or the conditions for their existence – are typically more demanding than for other social entities. Thomasson (2003) notes that many constructed social kinds and entities have no ‘deep’ application conditions that cannot be met merely through social agreement. But this is not the case for mathematical entities; new mathematical entities are not created by mere decision. Social agreement and intersubjective verification play a role in determining which mathematical results and entities are accepted as legitimate, but the construction of mathematical entities is also highly constrained in various other ways. My view is that there are multiple constraining factors that are present in cases of mathematical social construction in different combinations, but together the constraints ensure that the resulting entities are not arbitrary. Our mathematical practices and concepts are shaped by i) normative constraints, such as the goal of non-contradiction, logical principles, and a basic requirement of rational reasoning, ii) specific cognitive abilities as basis for mathematical knowledge, iii) systematic links between theories and different areas of mathematics, where important cases include definitions in terms of other entities and

existence proofs from accepted theories, and iv) applications, both within mathematics and in empirical sciences. (For similar accounts of constraints in mathematical practice, see Ferreirós 2016 and Wagner 2017.) The point is that whatever the relevant constraints are in each case, they impose further conditions that need to be met for mathematical entities to exist.

So, the reasons to deny magic in social ontology apply also to mathematics. But there is a further reason to take mathematical social constructions to be real, legitimate entities. The reason is that central parts of mathematics have close connections to certain features of the natural world (elementary arithmetic and geometry, in particular, and arguably also rudimentary logic). What the natural world is like and how we can interact with our environment constrains how we do mathematics. In the case of natural numbers and arithmetic, there are objective facts of how distinct physical objects can be manipulated and arranged. Additionally, we have innate cognitive abilities to detect quantities (subitizing and approximate number system), which form a universally shared basis for the social construction of natural numbers (see Pantsar 2021, 2024). Such connections to the natural world result in socially constructed entities that are more “objectively” real: mathematical entities are less up to us or culturally varied than most social entities, and they are more stable and robust than entities constructed in, for example, artistic practices.

In sum, social constructionism gives a lightweight ontology to mathematics that can avoid the threat of magic objections. Because abstract mathematical entities are social constructions, their existence does not require anything more mysterious than the appropriate practices. At the same time, the existence conditions of mathematical entities are more demanding than other social constructions due to the highly constrained nature of mathematical practices. This, together with connections to natural world, results in entities that are neither arbitrary, nor magic-like.

5. Conclusion: Mathematicians as artisans, not magicians

It should be noted that I am not alone in connecting magic arguments in mathematics and social ontology. In *Ontology Made Easy* (2015), Thomasson, too, talks of the easy existence of numbers and marriages alike, and makes arguments against metaphysical magic that apply both to mathematical and

social entities. However, my approach to the magic question regarding mathematical entities differs from that of Thomasson. Thomasson's view is that, while we can easily infer the existence of numbers and the like from uncontroversial conceptual truths, the entities exist quite independently of our language and concepts (2015: 217). Thus, there is no worry about things “popping into existence”, but rather:

In metaphysics as elsewhere, the only way to pull a rabbit out of a hat is if it's already there. (Thomasson 2015: 220)

In my social constructionist view, things are not quite as simple. Continuing with the metaphor, the first difference is that the rabbit was not already there, it had to be constructed out of suitable materials in a suitable context. For mathematical entities, the materials are something like concrete actions, symbols, and operations, and accepted concepts, definitions and methods, and the context consists of the historically situated social practices of mathematics. There is a close connection between the “stuff in the hat” and the rabbit, but they are not the same thing. (Elsewhere I suggest that the connection is a particular kind of grounding relation, but I leave this discussion for other occasions.)

Second difference is that I consider the person – or more appropriately, community – pulling the rabbit to have a significant role in the existence of mathematical entities. The thought is that without humans acting, talking, and thinking mathematically, abstract mathematical entities would not exist. Thus, my argument against metaphysical magic is not that mathematical entities do not need to be created, but that a better metaphor for the creation is that of a skilful artisan, not a magician. Socially constructing mathematical things like 4-dimensional cubes or massively large infinite sets is undoubtedly a very complex and rather weird kind of metaphysical artisanship. But the point of the comparison to ordinary social entities is that it is a kind of creation that humans do all the time, and not any kind of magic trick.

References

Cole, Julian C. (2015) “Social construction, mathematics, and the collective imposition of function onto reality”, *Erkenntnis*, 80(6), 1101–1124.

Effingham, Nikk (2010) “The metaphysics of groups”, *Philosophical Studies* 149, 21–67.

Feferman, S. (2014) "Logic, mathematics, and conceptual structuralism", in: Penelope Rush (ed.) *The Metaphysics of Logic*, Cambridge: Cambridge University Press, 72–92.

Ferreirós, José (2016) *Mathematical knowledge and the interplay of practices*, Princeton: Princeton University Press.

Hartimo, Mirja; Rytilä, Jenni (2023) "No Magic: From Phenomenology of Practice to Social Ontology of Mathematics", *Topoi* 42, 283–295.

Linnebo, Øystein (2018) *Thin objects*, Oxford: Oxford University Press.

Pantsar, Markus (2021) "Objectivity in Mathematics, Without Mathematical Objects", *Philosophia Mathematica* 29(3), 318–352.

Pantsar, Markus (2024) "Why do numbers exist? A psychologist constructivist account", *Inquiry: An Interdisciplinary Journal of Philosophy*. <https://doi.org/10.1080/0020174X.2024.2305386>

Rayo, Agustín (2015) "Nominalism, Trivialism, Logicism", *Philosophia Mathematica* 23 (1), nku013.

Rytilä, Jenni (2021) "Social constructivism in mathematics? The promise and shortcomings of Julian Cole's institutional account", *Synthese* 199, 11517–11540.

Searle, John R. (1995) *The construction of social reality*, London: Penguin.

Thomasson, Amie L. (2003) "Foundation for a Social Ontology", *ProtoSociology* 18, 269–290.

Thomasson, Amie L. (2015) *Ontology Made Easy*, Oxford: Oxford University Press.

Thomasson, Amie L. (2019) "The ontology of social groups", *Synthese* 196, 4829–4845.

Wagner, Roi (2017) *Making and breaking mathematical sense: histories and philosophies of mathematical practice*, Princeton: Princeton University Press.

Yablo, Stephen (2000) "A Priority and Existence" In Paul Boghossian and Christopher Peacocke (eds.) *New Essays on the a Priori*, Oxford: Oxford University Press, 197–228.

On the Nexus of Rule-Following, Learning, and Accountability in LLMs

Mirco Sambrotta (Bratislava, Slovakia)

Abstract

The rise of Large Language Models (LLMs), such as ChatGPTs, in contemporary generative AI marks a significant milestone in natural language processing, prompting deep inquiries into the core of linguistic comprehension within machine intelligence. This paper aims to assess the language understanding capabilities of LLMs through the theoretical lenses of Wittgenstein's and Brandom's frameworks. In light of Wittgenstein's seminal insights on rule-following and Brandom's inferentialism, to qualify as genuine language users, LLMs must exhibit the ability to follow inferential rules regardless of their internal states and processes. But this requires active participation in normative practices, which, in turn, implies the capacity to learn from social interaction. As a result, assessing LLMs' language understanding cannot solely rely on analyzing the texts they generate; it must encompass an examination of their language acquisition mechanisms. Considerations surrounding AI accountability and responsibility thus emerge as pivotal in shaping (also) the evolving dynamics and landscape of human-AI discursive interactions.

1. Introduction

The meaning of an expression is "its use in the language" (Wittgenstein, 1953: § 43). But what exactly is Wittgenstein getting at? He likens the meaning of an expression to its role within our language games: "Compare the meaning of a word with the "function" of an official. And "different meanings" with "different functions"" (Wittgenstein, 1969: 69).

Inferentialists add weight to this idea by suggesting that what sets apart conceptual contents is the role expressions assume within material consequence and incompatibility relations—in turn, propositional contents (which are a principal species of conceptual content) are what can perform the office both of premise and of conclusion in inferences (Brandom 1994; Peregrin 2014). In short, "It is only because the expressions in terms of which we describe objects... locate these objects in a space of implications, that they describe at all, rather than merely label" (Sellars 1958: 306-307).

Expressions acquire their content by taking part in the complex network of inferential relationships, which thus makes it possible for discursive practice—namely, the practice of explicitly claiming and judging that things are so-and-so. Such relationships equip the expressions with roles, "which may be considered as their "meanings"" (Peregrin 2021: 314). Grasping meanings (i.e., understanding) is then grasping the inferential roles expressions play. As a

result, being considered a competent language user hinges on the ability to navigate inference-making and drawing conclusions from given premises. That is, it necessitates the discernment of what constitutes evidence supporting a claim, as well as recognizing what is incompatible with it. In this way, mastery of a language ultimately boils down to a skill: to understand an expression, one must "know how" to use it.

2. LLMs and Rule-Following

Do Large Language Models (LLMs), such as GPTs and ChatGPTs, possess the aforementioned know-how? Undoubtedly, it can be argued that LLMs engage in inference-making, discern implications or contradictions between sentences, deduce conclusions from given premises, and map out reachable steps and excluded derivations. Should we take this as evidence to conclude that LLMs are genuine language users and understand language?

In order to answer this question, it is crucial to recognize that roles are conferred by rules. Indeed, Wittgenstein's own stand on language games suggests that they essentially operate as rule-based systems. Within these systems, expressions fulfill their specific functions dictated by the rules that govern their usage. This underscores that the significance of linguistic elements relies on the norms that regulate their application in judgment.

Brandom's (1994) "inferentialism" builds on this idea, proposing that the meaningfulness of an expression lies in its occurrence being subject to inferential norms (i.e., the norms underlying its material consequence and incompatibility relations). Understanding a word, hence grasping its meaning, consists in mastering those very rules. Thus, the relevant question to ask is: Do LLMs follow inferential rules?

Traditional computer systems rely on software states, which essentially encompass computational operations or processes (such as codes, algorithms, and software implementations). Software states can be understood as a device's internal states: its 'mental life'. Can we then attribute to machines the ability to follow rules in light of such internal states?

As Wittgenstein taught us, rule-following is not a matter of what is in the mind. For instance, delving into the internal representation a pupil consults when correctly continuing a series of numbers is a pointless (and perhaps impossible) endeavor (Wittgenstein 1953: §143-§185). Indeed, attempting to explain rule-governed activities by appealing to subjective "interpretation" at each instantiation of the rule inevitably results in infinite regress (Wittgenstein 1953: § 201). If the correct execution of an action each time necessitated consulting the guiding rule (i.e., the proposition), then the act itself of consulting that rule would in turn presuppose consulting the rule that guides the act of consulting the rule previously consulted, and so on. Rule-following does not hinge on internal states. Accordingly, it cannot be accounted for in terms of computation: If a computer's output relies on its internal state for meaning, that internal state cannot possess meaning without additional states, thus initiating an infinite regress, wherein each state requires further state for validation. To count as language users, human beings must exhibit the ability to follow the rules that govern language use regardless of their internal mental states. Likewise, to count as genuine language users, computers must exhibit the ability to follow the rules that govern language use regardless of their internal mechanisms, such as codes, algorithms, or computational processes of any kind.

Nevertheless, contrary to classical von Neumann architectures (i.e., the traditional computer systems), the neural network form has turned out to be able to address topics and solve problems without requiring explicit programming and formulation of rules. This might give rise to epistemic challenges because it becomes difficult to predict the behavior of autonomous devices or trace causal connections to individuals controlling the outcomes. That is to say, where behavioral output exceeds the purposes of initial programming, our efforts to decipher causes might not succeed. As a result, some neural networks may well remain "black boxes." The same happens in sufficiently large models, where it is possible to confirm the existence of features for which the models have not been directly programmed. In this case, we are then unable to interpret the inner workings of LLMs, and "any attempt at a precise explanation of an LLM's behavior is doomed to be too complex for any human to understand" (Bowman 2023). That's why, some LLMs too are frequently referred to as black boxes. Does that mean that we

should then conclude that LLMs are really following rules irrespective of their internal states and mechanisms?

3. The Normativity of Norms

Rules cannot but be a social matter! That is, rules must be understood as a public business; they are implicit in the social practice. But with social practices comes normativity. This means that norms are instituted socially: by "reciprocal recognition," by attributing to each other normative statuses (e.g., responsibility and authority). In turn, such normative statuses are instituted by adopting normative attitudes towards one another: the attitudes of taking themselves and others to be situated and bound up in a web, network, and constellation of rights and obligations. Norms come into existence only when there are beings that adopt to themselves and others such normative attitudes. Accordingly, talking of "roles" is then talking of the canons of authorization and prohibition sustained within the social community.

However, to say that social norms are instituted when we take or treat each other as responsible, authoritative, and so on, is to say that social norms are instituted when we take or treat performances as appropriate or inappropriate, correct or incorrect. And practically sanctioning or rewarding performances is one way of treating them as correct or incorrect. Then, by setting the standard for assessments of correctness, the norms one applies to determine what one has made oneself responsible for in applying them.

It is important to notice, though, that norms are not prior to practice. What explains the normativity of norms is the practice itself. Consequently, norms are not accessible from outside the practices that determine them. Norms are accessible only from within participating in the social practices in question, which determine their content. This means that to access the norms, one has to attribute normative statuses to oneself and others, and one does so by adopting normative attitudes towards oneself and others. Ultimately, one accesses the norms by contributing to shaping them and determining what they are.

This restriction would also apply to "God" (Esfeld and Köstner 2022: 12). Such an omniscient being would know all the facts about the world. Nonetheless, He

(?) would have to participate in practices to know the norms that are determined in them, and by participating, He would contribute to determining those very norms. This is a consequence of these norms not being accessible from outside the practices that determine them. That is because normative statuses (and hence the normativity of norms) only show up as in social practice, and there is no perspective from nowhere available to access such practices.

4. Discursive Norms

Drawing on a view taken from Hegel by the American pragmatists (C. S. Peirce, William James, and John Dewey), and later on by Wittgenstein, norms governing linguistic performances are to be understood as implicit in social practices. As such, they are instituted only when beings adopt to themselves and others normative attitudes.

Brandom (1994: Ch. 1) elaborates this view in terms of the normative attitudes of commitment, entitlement (and precluded entitlement) to commitment. Making a claim involves undertaking a commitment to “demonstrate one’s entitlement to the claim, if that entitlement is brought into question,” where being entitled to a claim is being “entitled to make it” (Brandom 1994: 171-172). This roughly means that one is making a claim insofar as “He is making a move in a practice of giving and asking for reasons, in which one move has normative consequences for what others are obligatory, permitted, or prohibited” (Brandom 2014: 354). Normativity is thus presupposed for the very formulation of claims: claims imply a normative space where they can be criticized and justified, the normative web of giving and asking for reasons, which is constituted exactly by those very social-normative relations. Accordingly, “To understand an assertional speech act is to know how to keep score on the commitments the speaker has undertaken by performing that act” (Brandom 2014, 357). That is, understanding a claim is taking up a stance in a network of related possible claims, which stand to one another in rational relations of material consequence and incompatibility. It follows that talking of “linguistic roles,” roles in practices of giving and asking for reasons, is talking of the canons of commitment and entitlement, sustained within the speech community.

We take each other to express meanings when hold each other committed and entitled. The relevant normative practice is therefore conceived as constitutive of conceptual and propositional contents, and “basic in the order of semantic explanation” (1994: 496; 1983: 640, 642, 644). Meaningfulness is built into such practice, within which alone meanings get expressed! It emerges from the practice that has the right normative and social structure (Brandom 1994: Ch. 3), wherein individuals adopt instituting normative attitudes to one another and, in turn, attribute normative statuses to each other. This view clearly stems from the work of Wittgenstein, according to which, the capacity to make propositionally explicit claims and have conceptually contentful thoughts is intelligible only in the context of implicitly normative social linguistic practices (1953: §199).

However, again, to say that discursive norms are instituted when we take or treat each other as committed and entitled is to say that social norms are instituted when we take or treat performances as appropriate or inappropriate, correct or incorrect. And, again, practically sanctioning or rewarding performances is one way of treating them as correct or incorrect. This is, norms that confer inferential roles (which constitute meaning) are set up by what is fixed as correct or incorrect inferences in social interactions: “... inferences are correct in the sense that they are accepted in the practice of a community,” where communal acceptance is a matter of “actual practical attitudes” of assessment (Brandom 1994: 137).

5. Learning and Language Acquisition

Pursuing this order of explanation leads to the conclusion that exercising “a certain form of” authority and taking “a certain kind of” responsibility is exactly what one must do in order to count as grasping and understanding what, in social practice, shows up as conceptual and propositional contents, and thus to count as a competent language user. In particular, “To be a competent speaker (competent concept user) is to engage in practices of giving and asking for reasons, that is, making inferentially articulated assertions and justifying them. To do so, one must attribute and acknowledge commitments and entitlements, and practically keep track of their inferential relations” (Brandom 2014: 359-360).

But can LLMs be held accountable in a normative sense for their output? Can we adopt normative attitudes towards them? In other words, can we engage with them through what Dennett (1987) calls "the intentional stance" and Sellars (1962) characterizes as "rehearsing an intention"? What is needed for us to attribute them responsibility is to be in principle possible for us to assess their behaviour (i.e., their attempts at rule-following) as right or wrong. In turn, what is needed for LLMs to count as responsible is to be in principle possible for them to improve in response to (output) failures, hence to learn from their mistakes.

As Ryle (1949) pointed out, only abilities that become manifest as "acquired" count as knowledge-how. Thus, it is required that we learn those abilities (vs. innate abilities) to count as knowing how to exercise them (vs. mere habits). Intelligence manifests itself in learning! However, it is only by reciprocating normative attitudes that learning can take place. Teacher-student relation obtains when the student does something wrong (i.e., she does not follow the teacher's instructions), and the teacher asserts power in response to it: "If a child does not respond to the suggestive gesture, it is separated from the others and treated as a lunatic" (Wittgenstein 1958: §30). Instead of the tool/user model, we have here a complex relationship between the teacher and the student, in which both share responsibility and authority. We praise or blame the skill and sacrifice of the teacher and laud or sanction the actions of the student as well.

Language is always learned in this way: one learns how to follow the rules of language in practice, through trial and error. In turn, language is always taught by treating performances as correct or incorrect by practically sanctioning or rewarding them. Ultimately, the social practices that allow the infant to acquire their first language are therefore of the same kind as the social practices that imbue strings of letters with meaning (and reference).

6. Conclusion

According to the view put forward here, the claim that LLMs of generative AI are language users cannot be based solely on the indistinguishability of the generated texts from human texts, but should rather be based on the similarity of their language acquisition mechanisms to our own mechanisms of language

acquisition. LLMs can be expected to follow rules that govern the use of linguistic locutions (i.e., rules of inferences), and so mastering the use of words and grasping and understanding their meaning, only if they have learned to do so in practice through social sanction and reward. But do LLMs learn in this way? Do we practically train them?

The current way in which language models are trained involves detecting statistical patterns: they capture statistical distributions of sequences of words in a collection of texts, so generating new texts drawn from the same distribution. LLMs are trained to predict words from massive datasets of text from the internet. If so, they are just sophisticated implementer of “surface statistics” (Li 2023); mere repeaters guided only by probability. They merely assemble words into sentences based on statistical and probabilistic information about how to combine them, and thus they behave like a kind of “stochastic parrots” (Bender et al. 2021). That is, the prevailing view is that language models only operate stochastically with linguistic expressions, generating text according to pre-set stochastic patterns stored. We can then compare the performance of LLMs to, for example, the jargon-spouting students who try to imitate their professors but basically do not know what they are talking about (Browning and Lecun 2022). As a result, maybe, we are more inclined to view machine language performance as a mere simulation of understanding (e.g., a mere syntactic manipulation), rather than to see it as based on a proper understanding of language. Indeed, it would perhaps be naive to think that one is guided by these statistics probabilistic repetition of patterns in learning language.

Therefore, although the future holds promise for assessing AI as genuine language users, the current landscape still presents considerable challenges in realizing this vision.

References

Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina; Shmitchell, Shmargaret (2021) “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23, Virtual Event Canada: ACM.

Bowman, Samuel R. (2023) “Eight Things to Know about Large Language Models”, *arXiv*. <https://doi.org/10.48550/arXiv.2304.00612>

Brandom, Robert (1983) "Asserting", *Noûs* 17 (4), 637-650.

Brandom, Robert (1994) *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, Mass.: Harvard University Press.

Brandom, Robert (2014) "Intentionality and language", in: Enfield NJ, Kockelman P, Sidnell J, (eds.) *The Cambridge Handbook of Linguistic Anthropology*, Cambridge University Press, 347-363.

Browning, Jacob; Lecun, Yann (2022) "AI and The Limits of Language", *Noema*. <https://www.noemamag.com/ai-and-the-limits-of-language>

Dennett, Daniel C. (1987) *The intentional stance*, Cambridge (Massachusetts): MIT Press.

Esfeld, M. and G. Köstner (2022) "Normative relations, mind points and social ontology", *Synthese* 200 (6),1-15.

Li, Kenneth (2023) "Do Large Language Models learn world models or just surface statistics?", *The Gradient*. <https://thegradient.pub/othello/>

Peregrin, Jaroslav (2014) *Inferentialism: Why Rules Matter*, New York: Palgrave Macmillan.

Peregrin, Jaroslav (2021) "Do Computers 'Have Syntax, But No Semantics'?", *Minds and Machines* 31, 305–321.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Sellars, Wilfrid (1958) "Counterfactuals, Dispositions, and Causal Modalities" in: Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.) *Minnesota, Studies in the Philosophy of Science, Volume II: Concepts, Theories, and the Mind-Body Problem*, Minneapolis: University of Minnesota Press, 225-308.

Sellars, Wilfrid (1962) "Naming and Saying", *Philosophy of Science* 29 (1), 7-26.

Wittgenstein, Ludwig (1953) *Philosophical Investigations*, Oxford: Blackwell.

Wittgenstein, Ludwig (1958) *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigations'*, Oxford, England: Harper & Row.

Wittgenstein, Ludwig (1969) *On Certainty*, Oxford: Blackwell.

Acknowledgement

This work was supported by the Slovak Research and Development Agency, under Contract No. APVV-22-0323 (project "Philosophical and methodological challenges of intelligent technologies"), the John Templeton Foundation for the project "The Necessity of God's Existence: Divine Names and Modality" (Ref. R64145/CN053), and the Štefan Schwarz Support Fund for the grant No. 2021/OV3/017.

Die Papierkrone am Schachkönig – zur Metaphorik von Wittgensteins Sprachkritik

Alfred Schmidt (Vienna, Austria)

Abstract

Wittgenstein verwendet zur Charakterisierung seiner sprachkritischen Methode bzw. seines Verständnisses von Bedeutung eine Reihe anschaulicher Metaphern. Dazu gehören das Bild von der Papierkrone am Schachkönig, die Analogie von den leerlaufenden Räder und vom Feiern der Sprache, wie auch die Metapher vom Fluss des Lebens, der sprachlichen Ausrücken erst eine bestimmte Bedeutung gibt. Ziel des Vortrages ist es, die Aussagekraft dieser Analogien und ihr Verhältnis zu einander genauer zu analysieren.

Die Faszination bzw. Lebendigkeit von Wittgensteins Philosophie beruht zu einem nicht geringen Teil auf seiner Verwendung origineller, anschaulicher Bilder, bzw. Metaphern. Die unmittelbare Evidenz dieser Analogien tritt dabei häufig an die Stelle eines linearen rationalen Argumentierens. Es gelingt ihm damit komplexe begriffliche Zusammenhänge intuitiv verständlich zu machen, bzw. seine Leser*innen aus eingefahrenen Denkgewohnheiten herauszureißen. Umgekehrt war sich Wittgenstein aber auch der verführerischen Kraft dieser Bilder bewusst (PI 2009: § 115 “Ein Bild hielt uns gefangen...”), die stets einer rationalen “Übersetzung” bzw. Prüfung zugänglich bleiben müssen.

In meinem Vortrag geht es um den Vergleich einiger Metaphern oder Bilder, die Wittgenstein verwendet, um seine Kritik an einem falschen Verständnis von Bedeutung zum Ausdruck zu bringen.

1. Die Papierkrone am Schachkönig

Eine der zahlreichen anschaulichen Bilder, die Wittgenstein verwendet, um den Unterschied von bedeutungsvoller und bedeutungsloser Sprache zum Ausdruck zu bringen, ist das Bild von der Papierkrone am Schachkönig im *Blue Book*:

– I want to play chess, and a man gives the white king a paper crown, leaving the use of the piece unaltered, but telling me that the crown has a

meaning to him in the game, which he can't express by rules. I say: “as long as it doesn't alter the use of the piece, it hasn't what I call a meaning.
(DIC 309: 110)

Das Schachspiel war in dieser Periode – von etwa 1930 bis Mitte der 30er Jahre – für Wittgenstein eines der zentralen Paradigmen zum Verständnis von Sprache, wobei er aber gleichzeitig auch auf die Grenzen dieser Analogie aufmerksam machte (vgl. dazu Schmidt 2015). Das etwas bizarre Verhalten des Schachspielers, hat keine Bedeutung für das Spiel, weil es auf die Verwendung der Schachfigur keinen Einfluss hat, meint Wittgenstein. Hätte sie einen, d.h. würde die Papierkrone den Schachkönig zu einer anderen Figur machen, so würde das von einem Schachschiedsrichter wohl kaum toleriert werden. Es wäre nicht mehr das, was wir Schachspiel nennen. Das Schachspiel beruht auf der durch allgemein gültige Regeln geleiteten Verwendung bzw. Bewegung von Schachfiguren auf einem Schachbrett.

Wofür die Analogie der Papierkrone am Schachkönig steht, erklärt Wittgenstein an der gleichen Stelle:

If someone taught me the word “bench” and said that he sometimes or always put a stroke over it thus: “bench”, and that this meant something to him, I should say: “I don't know what sort of idea you associate with this stroke, **but it doesn't interest me** unless you show me that there is a use for the stroke in the kind of calculus in which I wish to use the word “bench”. (DIC 309: 110)

Der Strich über dem Wort ‘bench’ entspricht der Papierkrone, beides gehört in den Bereich der bedeutungslosen Ornamentik. Sie stehen für private Erlebnisse, die das Sprechen bzw. Schachspielen begleiten und sind damit dem ominösen Käfer in der Schachtel (PI 2009: §293) verwandt, den nur der jeweilige Besitzer sehen kann, weil nur er die Schachtel öffnen kann.

Der Käfer entspricht den psychischen Zuständen des Schachspielers während der Partie. Wenn er uns seine Erlebnisse während Partie nachher zu erklären versucht: „Diesen Zug machte ich mit der Absicht, In dieser Stellung fühlte ich mich sehr überlegen, usf.“, so lässt sich mit Wittgenstein darauf antworten:

Deine psychischen Erlebnisse während der Schachpartie interessieren uns hier nicht. Für das Spiel relevant sind ausschließlich die Züge am Schachbrett, sie entscheiden über Gewinn oder Verlust.

Versucht man die Bedeutung sprachlicher Ausdrücke durch private Bedeutungserlebnisse zu erklären, die dem anderen Gesprächspartner notwendiger Weise verborgen bleiben, so steckt man genau in jenem irreführenden Bild von Bedeutung bzw. Sprache fest, das Wittgenstein bekämpft.

Ist die Papierkrone am Schachkönig aber deswegen wirklich völlig bedeutungslos? Müssen wir das Beharren des Spielers, dass die Papierkrone **für ihn** sehr wohl eine Bedeutung habe, nicht dennoch ernst nehmen? Er könnte z.B. behaupten: "Die Papierkrone auf meinem König hilft mir die richtigen Züge zu finden." Und das könnte sehr wohl auch an seinen Spielergebnissen objektiv ablesbar sein: d.h. er gewinnt immer dann, oder zumindest wesentlich öfter, wenn er die Papierkrone verwendet. Wittgenstein würde einen solchen Zusammenhang wohl nicht bestreiten, aber sagen, diese Phänomene interessieren uns hier nicht. Wenn wir eine Schachpartie analysieren, wollen wir wissen, welche Züge gut oder schlecht waren, in welcher Situation es bessere gegeben hätte usf., nicht aber, **warum** die Spieler so und so gezogen haben. Für einen Schach-Psychologen hingegen mag gerade das die ihn interessierend Fragestellung sein.

2. Leerlaufende Räder

Ein der Papierkrone am Schachkönig eng verwandtes, von Wittgenstein häufig verwendetes Bild ist das der leerlaufenden Räder, etwa in PU 132: „Die Verwirrungen, die uns beschäftigen entstehen gleichsam, wenn die Sprache leerläuft, nicht wenn sie arbeitet.“ Die Sprache läuft dann leer, wenn wir die „philosophische“ Verwendung von Wörtern und Sätzen in keine konkreten Sprachspiele einordnen können (Pi 2009: 96: „Wozu sind diese Wörter nun zu gebrauchen. Es fehlt das Sprachspiel, worin sie zu verwenden sind“).

Wittgenstein exemplifiziert dies ausführlich etwa an seiner Kritik an G.E. Moore in seinen letzten Bemerkungen *Über Gewissheit*. Er bezieht sich darin wiederholt auf Moores Beispiele für unmittelbar evidente, unbezweifelbare

Sätze, wie etwa: ‚I know that that’s a tree.‘ . Für Wittgenstein ein typisches Beispiel eines leerlaufenden Rades, denn wir wissen nicht, welche konkrete Gesprächssituation wir uns zu diesem Satz vorstellen sollen:

"I know that that’s a tree." Warum kommt mir vor ich verstehe den Satz nicht? Obwohl er doch ein höchst einfacher Satz der gewöhnlichsten Art ist? Sowie ich aus der philosophischen an eine alltägliche Anwendung des Satzes denke, wird sein Sinn klar und gewöhnlich. (ÜG 197: 347)

Wittgenstein gibt sich Mühe, sich solche alltägliche Anwendungen vorzustellen, in denen dieser Satz sofort einen klaren Sinn bekommt, etwa :

Ich schaue auf eine Pflanze, die ich für eine junge Buche der Andere für eine Ribiselpflanze hält. Er sagt „Das ist ein Strauch“, ich, es sei ein Baum. – Wir sehen im Nebel etwas, was einer für einen Menschen hält, der Andre sagt „Ich weiß dass das ein Baum ist. (ÜG 1977: 349)

Sobald wir uns aber in dieser Weise eine konkrete Anwendung von Moores angeblich absolut gewissen Sätzen vorstellen, verschwindet ihre Unklarheit, aber damit auch ihre unzweifelhafte Gewissheit. Denn es könnte ja doch eine Ribiselstaude sein, oder ein Mensch im Nebel, was wir für einen Baum halten. Und genau das ist es, was Wittgenstein uns damit zeigen will.

Der springende Punkt im Bild des leerlaufenden Rades ist aber folgender:

Ich habe nichts dagegen, daß du an der Maschine der Sprache ein leerlaufendes Rad anbringst, aber ich wünsche zu wissen, ob es leer läuft oder in welche andere Räder es eingreift. (DIC 302: 29,30)

Leerlaufende Räder stören die Funktion der Maschine im Grunde nicht, also ist nichts dagegen einzuwenden, ebenso wenig wie gegen die Papierkrone am Schachkönig. Wittgensteins Argument ist aber, dass wir oft auf den ersten Blick nicht erkennen können, ob ein Rad leertläuft oder ob es nicht doch in den Lauf der Maschine eingreift. Wir sind versucht, wunderbar blinkende („metaphysische“) Räder anzustaunen, die aber doch mit der Funktion der Maschine (Sprache) nichts zu tun haben. Philosophische Verwirrungen haben genau diesen Ursprung meint Wittgenstein. Sein sehr pointierter Ausdruck:

“Ich wünsche zu wissen...” im obigen Zitat ist die Formulierung eines philosophischen Programmes bzw. von Wittgensteins philosophischer Methode.

Gibt es eine eindeutige Unterscheidung zwischen leerlaufenden und nicht leerlaufenden Räder aber überhaupt? Beim Beispiel der Papierkrone am Schachkönig haben wir gesehen, dass das scheinbar sinnlose Ornament sehr wohl einen Einfluss auf das Spiel haben kann, nur eben nicht auf der Ebene der Spielregeln, sondern auf der der Psyche des Spielers. Wittgenstein sieht in der Analogie der leerlaufenden Räder wohl eine scharfe Grenze: entweder das Rad greift in den übrigen Mechanismus der Maschine ein, oder nicht: Als ausgebildeter Ingenieur fällt es ihm nicht schwer, dies zu unterscheiden.

3. Vom Feiern und Arbeiten der Sprache

Ein weitere Metapher, die zum gleichen Themenkreis gehört, ist die Rede vom „Feiern der Sprache“. Wittgenstein stellt die Verbindung zu Leerlaufen explizit her: „Die Konfusionen die uns beschäftigen entstehen, gleichsam, wenn die Sprache feiert, nicht wenn sie arbeitet. (Man könnte sagen: ‚wenn sie leerläuft‘)“. (PI 2009: §38).

Das Feiern bezieht sich dabei nicht alleine auf die Sprache, wie Wittgenstein an anderer Stelle erläutert:

Man kann sagen: wenn wir Philosophieren feiert nicht nur unsre Sprache, sondern auch unser Blick. Denn während ich den Ofen heize, sehe ich ihn anders als, wenn ich beim Philosophieren auf ihn starre. || , denke ich nicht an den ‘visuellen Ofen’, das Sinnesdatum, etc. (MS 120: 49)

Mit der Metapher vom Feiern der Sprache deutet Wittgenstein an, dass zum Leerlauf – dem Sprechen außerhalb konkreter Sprachspiele – auch noch eine besondere Feierlichkeit oder angenommene tiefere Bedeutsamkeit hinzukommt, die den metaphysischen Sprachgebrauch meist begleitet. Wittgensteins negative Einstellung zum Feiern ist wohl auch als Teil einer protestantischen Arbeitsethik zu sehen, die er aus seinem Elternhaus mitbekommen hatte. Menschen werden an ihren Leistungen gemessen, feiern und Müßiggang sind ein Zeichen von Faulheit. Bezeichnend ist etwa folgende

Episode, aus den Erinnerungen von Maurice O'Connor Drury während Wittgensteins Aufenthalt in Rosro zusammen mit Francis Skinner und Drury im Sommer 1935:

Als wir schließlich den Sandstrand sahen, erblickten wir unter uns die Familien Mortimer – die einzigen Bewohner dieser einsamen Gegend - , die gerade dabei waren, auf dem winzigen kultivierbaren Landstrich, der ihnen zur Verfügung stand, Heu zu ernsten. Sobald Wittgenstein dies sah, wandte er sich um:

W.: Wir gehen jetzt zurück. Diese Leute dort sind bei der Arbeit, und es ist nicht richtig, das wir vor ihren Augen Ferien machen. (Drury 1987: 181)

Als Müßiggänger schämt man sich vor den arbeitenden Menschen. Ebenso beschämend ist für Wittgenstein der müßige Sprachgebrauch der Philosophen. Wie Sprache arbeitet, hat er paradigmatisch in den elementaren Sprachspielen vom Einkaufen im Krämerladen (PI 2009: 1) bzw. der Baustelle (PI 2009: 2) demonstriert. In PU 23 zeichnet Wittgenstein ein breites Bild der Mannigfaltigkeit der Sprachspiele, in denen Sprache auf je eigene Weise arbeitet:

- Befehlen, und nach Befehlen handeln –
- Beschreiben eines Gegenstands nach dem Ansehen, oder nach Messungen
-
- Herstellen eines Gegenstands nach einer Beschreibung (Zeichnung) –
- Berichten eines Hergangs –
- Über den Hergang Vermutungen anstellen –
- Eine Hypothese aufstellen und prüfen –
- Darstellen der Ergebnisse eines Experiments durch Tabellen und Diagramme –
- Eine Geschichte erfinden; und lesen –
- Theater spielen –
- Reigen singen –
- Rätsel raten –
- Einen Witz machen; erzählen –
- Ein angewandtes Rechenexempel lösen –
- Aus einer Sprache in die andere übersetzen –

Bitten, Danken, Fluchen, Grüßen, Beten.
(PI 2009: §23)

Philosophieren kommt in Wittgensteins Aufzählung nicht vor.

Warum sollte die Sprache aber nicht auch feiern dürfen? Gilt hier nicht das Gleiche wie bei den leerlaufenden Rädern? D.h. Feiern ist in Ordnung, solange es nicht als Arbeit ausgegeben wird. Feiert die Sprache wie z.B. in der Poesie, insbesondere in der Lyrik, so ist ihr Feiern klar erkenn- und daher tolerierbar.

4. Im „Fluß des Lebens“

Wittgensteins vielleicht schönste und allgemeinste Metapher, die seine sprachkritische Methode zum Ausdruck bringt, ist die vom ‚Fluß des Lebens‘: „Denn die Wörter haben eben nur im Fluß des Lebens Bedeutung.“ (MS 137: 66a). Norman Malcolm schreibt in seinen Erinnerungen, diese Äußerung Wittgensteins erschien ihm damals wie heute als Resümee eines größten Teiles seiner Philosophie (Malcom 1987, 122). Wittgenstein variiert dieses Bild an einigen weiteren Stellen:

Ich möchte sagen: das Gespräch, die Anwendung & Ausdeutung der Worte fließt dahin, & nur in diesem || dem Fluß || nur im Fluß hat das Wort seine Bedeutung.“
(Ms 131: 22)

Die Worte stehen in einem Fluß. Nur in einem Leben haben diese Worte || sie ihren Sinn. (MS 137: 41)

Auch was im Inneren vorgeht hat nur im Fluß des Lebens Bedeutung. (MS 169: 47v)

Erst der ‚Fluss des Lebens‘, d.h. ihr Einbindung in konkrete Kontexte, Sprachspiele, gibt Sätzen eine klare Bedeutung. Metaphysik – so könnte man in diesem Bild bleibend sagen – ‚staut‘ den natürlichen Fluss des Lebens und der Sprachverwendung. So entsteht zwar der Anschein von Tiefe, aber es ist trübes, schlammiges Wasser, in dem Ausdrücke ihren klaren Sinn verlieren. Maurice Drury berichtet, dass Wittgenstein geradezu empört war über einen philosophischen Missbrauch von Wörtern, der sie gleichsam in einen

„academic isolation ward“ führt, abgetrennt vom „stream of life“ ihrer alltäglichen Verwendung:

He told his class, that this is the very last thing philosophy should do ... it is shocking to use words with a meaning they never have in normal life and is the source of much confusion. (Drury 2019: 8)

Wittgensteins scheinbar grenzenloses Vertrauen in den Strom des Lebens, der alle Verunreinigungen und philosophischen Verwirrungen in unseres Sprachgebrauch beseitigen kann, mag vielleicht den Anschein von Naivität erwecken. Stanley Cavell schreibt:

Ich nähere mich den Untersuchungen direkt, ohne Umweg über den Tractatus, und dabei erstaunt mich am meisten diejenige Seite von Wittgensteins Denken, die der Alltagssprache vertraut und in der Kraft des Alltäglichen Ruhe gegenüber den tiefen Beunruhigungen unserer philosophischen Mißdeutungen findet. ... Die Kraft dieser Anerkennung des Alltäglichen für die Philosophie hängt untrennbar mit der Einsicht zusammen, daß in der Zurückweisung oder Verbiegung der Ordnung des Alltäglichen ein Grund für philosophische Leere (sprich Nichtigkeit) und Gewalt liegt.

(Cavell 2001: 7)

Gegenüber diesem ‚Urvertrauen‘ Wittgensteins in den tatsächlichen, alltäglichen Sprachgebrauch (PI 2009: §124 „Die Philosophie darf den tatsächlichen Sprachgebrauch in keiner Weise antasten...“) zeigt sich mitunter aber auch eine Spannung und ein Misstrauen, das Wittgenstein an überraschenden Stellen wie etwa dieser ausdrückt:

Die Menschen sind tief in den philosophischen i.e. grammatischen Konfusionen eingebettet. || Und sie daraus zu befreien setzt voraus, daß man sie aus den ungeheuer mannigfachen Verbindungen herausreißt in denen sie gefangen sind. Man muß sozusagen ihre ganze Sprache umgruppieren. (MS 113: 23)

Die Unantastbarkeit des tatsächlichen Sprachgebrauchs steht in einem scheinbar eklatanten Widerspruch zur Notwendigkeit die ganze Sprache umzugruppieren. Gerade aus diesen unaufgelösten Gegensätzen erwächst die anhaltende Faszination von Wittgensteins Philosophie.

References

Cavell, Stanley (2001) "Wittgenstein als Philosoph der Kultur. Alltäglichkeit als Heimat", in: Stanley Cavell, *Nach der Philosophie. Essays*, Berlin: Akad.-Verlag.

Drury, Maurice O'Connor (1987) "Gespräche mit Wittgenstein", in: Rhees, Rush (1987): *Ludwig Wittgenstein. Porträts und Gespräche*, Frankfurt: Suhrkamp, 142-235.

Drury, Maurice O'Connor (2019) *The Selected Writings of Maurice O'Connor Drury. On Wittgenstein, Philosophy, Religion, and Psychiatry*. Edited and introduced by John Hayes. London: Bloomsbury.

Kross, Matthias (2010) "Metapher als Entzug – Metapher als Gabe: Wittgenstein mit Blumenberg lesen?", in: *Essays on the philosophy of Wittgenstein*, herausgegeben von: Volker Munz, Berlin: DeGruyter, 75-92.

Malcom, Norman (1987) *Erinnerungen an Wittgenstein*, Frankfurt: Suhrkamp.

Schmidt, Alfred (2015) "Wittgensteins Schachparadigma – oder: Beruht unsere Sprache auf Regeln?"; in: *Realismus – Relativismus – Konstruktivismus. 38. Internationales Wittgenstein Symposium, 9. – 15. August 2015, Kirchberg am Wechsel*, Hrsg: Christian Kanzian, Josef Mitterer, Katharina Neges; Kirchberg: Österreichische Ludwig Wittgenstein Gesellschaft, 264-267.

Wittgensteins logischer Atomismus und Sokrates' Traum

Jan Schnalke (Cologne, Germany)

Abstract

Der logische Atomismus postuliert die Existenz unanalysierbarer Elemente, die durch einfache Zeichen repräsentiert werden und die Grundbausteine für wahrheitsfähige Sätze bilden. In dieser Untersuchung wird die Position des logischen Atomismus näher beleuchtet, indem gezeigt wird, dass sich ihre grundlegenden Annahmen bereits in Platons *Theaitetos* finden lassen und später von Wittgenstein weiter ausgebaut werden.

Für den logischen Atomismus spricht, dass durch das Postulat logischer Atome gewährleistet wird, dass wahrheitsfähige Sätze in der Wirklichkeit fundiert sind. Dennoch lässt sich durch Sokrates' Kritik im *Theaitetos* die Annahme logischer Atome problematisieren. Auch Wittgenstein argumentiert in den *Philosophischen Untersuchungen* im Geist des *Theaitetos* gegen die Position des logischen Atomismus. Er argumentiert, dass Sätze eben nicht auf immer fundamentalere Sätze reduzierbar sind, die letztlich aus logischen Atomen bestehen. Darüber hinaus lehnt er die Existenz einfacher Objekte, d.h. logischer Atome, im absoluten Sinne ab. Gemäß dem späten Wittgenstein ist die Einfachheit oder Komplexität eines Objekts kontextabhängig. Diese Revision zeigt einen Wandel in Wittgensteins Denken von der früheren Theorie logischer Atome hin zu einem kontextabhängigen Verständnis von Sprache und Bedeutung.

1. Einleitung

Grundannahme der vorliegenden Arbeit ist, dass Platons *Theaitetos*, Wittgensteins *Tractatus Logico-Philosophicus* und seine *Philosophischen Untersuchungen* die Idee des *Logischen Atomismus* behandeln. Der Begriff *Logischer Atomismus* geht wohl auf Russel zurück (vgl. Proops 2017), wobei dieser jedoch gesteht, die Grundidee des logischen Atomismus stamme von Wittgenstein (vgl. Russel 2010: 1). Eigentlich finden sich die ersten Züge für einem logischen Atomismus aber bereits in Platons *Theaitetos*. Dort wird er Gegenstand der Kritik Sokrates. Wittgenstein dagegen verteidigt die Idee des logischen Atomismus im *Tractatus* zunächst, auch wenn er in den *Philosophischen Untersuchungen* seine Haltung gegenüber diesem revidieren wird.

Ziel dieser Arbeit ist es die Bezüge zwischen diesen drei Texten herauszustellen. Dafür wird zunächst herausgearbeitet, inwiefern sich die Idee *Logischen Atomismus* bereits im *Theaitetos* finden lässt und sie dem Ansatz des *Tractatus* entspricht. Danach wird Wittgensteins Argument für die Plausibilität des logischen Atomismus erläutert, um anschließend Sokrates Kritik am

logischen Atomismus geltend zu machen. Im nächsten Schritt wird Wittgensteins Revision dargestellt. Zum Schluss werden die Ergebnisse zusammengefasst.

2. Der Logische Atomismus in Platons *Theaitetos*

Beim logischen Atomismus handelt es sich um eine Theorie über die Zusammensetzung von wahrheitsfähigen Sätzen (vgl. Ryle 1990: 31). Eine Theorie kann dann als *Logischer Atomismus* bezeichnet werden, wenn sie die Kombination der folgenden Thesen beinhaltet: (1) Bei der Analyse von wahrheitsfähigen Sätzen stößt man letztlich auf fundamentale Sätze, die sich nicht weiter analysieren lassen. (2) Diese fundamentalen Sätze bestehen aus semantisch einfachen Symbolen, d.h. *Namen*. (3) Diese *Namen* beziehen sich auf einfache Gegenstände, logische Atome, die sich nicht weiter zergliedern lassen (vgl. Proops 2017).

Gilbert Ryle hält Platons *Theaitetos* für eine Inspirationsquelle des logischen Atomismus des 20. Jahrhunderts (vgl. Ryle 1990). Im *Theaitetos* referiert Sokrates einen Traum, eine Theorie, die er gehört hat (vgl. Tht. 1992: 201e). Dieser Traum handelt von „Urbestandteilen“, aus denen alles Übrige zusammengesetzt sei. Die Urbestandteile selbst ließen aber keine „Erklärung“ zu. Somit könne man von ihnen weder sagen, dass sie *sind*, noch dass sie *nicht sind*. Auch könne man von ihnen nicht sagen, dass sie sie „selbst“ sind, oder dass sie „dieses“ oder „jenes“ sind. Die Urbestandteile lassen sich dem Traum zufolge nur benennen. Allererst durch ihre Verbindung, also durch die „Verflechtung“ von Namen, die sich auf die Urbestandteile beziehen, entstünden Erklärungen (vgl. Tht. 1992: 201e-202c).

Die Urelemente des Traumes lassen sich auch als logische Atome bezeichnen. Die logischen Atome werden mit Namen bezeichnet und aus ihrer Verbindung ergeben sich wahrheitsfähige Sätze. Der Theorie des Traumes zufolge können sich *nur* aus der Verbindung von Namen wahrheitsfähige Sätze ergeben. Ein Name eines logischen Atoms allein kann dagegen nicht wahrheitsfähig sein. Dies leuchtet insofern ein, als eine sprachliche Äußerung mindestens aus einem vollständigen Satz bestehen muss, um wahrheitsfähig zu sein. Es wäre höchst kontraintuitiv, zu behaupten, die bloße Äußerung „Fisch“ könne wahr oder falsch sein (vgl. Gabriel 2012: 49-50). Namen allein können also weder

wahr noch falsch sein und damit auch nicht geglaubt oder gewusst werden. Sie können eben nur benannt werden.

Sokrates Behauptung, dass logische Atome nicht sie „selbst“, „dieses“ oder „jenes“ sein können, deutet darauf hin, dass logische Atome nicht in Zusammenhänge eingebettet werden können, ohne ihren Status als logische Atome zu verlieren. In einer Aussage wie „dies ist ein Mann“ verbindet man bereits einen Namen zu einem wahrheitsfähigen Satz und geht nicht auf das logische Atom allein ein, das der Name bezeichnet. (vgl. Gabriel 2012: 34). Außerdem werden logische Atome als einfach vorgestellt. Damit ist gemeint, dass die Urbestandteile, auf die sich die Namen beziehen, sich nicht weiter analysieren lassen. Zum Beispiel kann der Name *Junggeselle* sich nicht auf ein logisches Atom beziehen, da sich dieser weiter in *unverheirateter Mann* zergliedern lässt. Einfache Namen wären also nur solche, die sich auf logische Atome beziehen, die man am Ende einer solchen Analyse erreicht (vgl. Ryle 1990: 30-31).

3. Der Logische Atomismus im *Tractatus*

Von einer solchen Theoriekonzeption, die Sokrates im *Theaitetos* referiert, lassen sich Parallelen zur Position Wittgensteins ziehen, die er im *Tractatus* vertritt. Wittgenstein selbst merkt in seinen *Philosophischen Untersuchungen* an, dass es sich bei den „Urbestandteilen“ des *Theaitetos* um die „Gegenstände“ des *Tractatus* handelt (vgl. PU 2022: § 46).

Von den Gegenständen des *Tractatus* behauptet Wittgenstein, sie seien „einfach“ (vgl. TLP 1922: § 2.02) und dass sie nur benannt werden können (vgl. TLP 1922: § 3.221). Diese Gegenstände gehen mit anderen Gegenständen Verbindungen ein. Die Verbindungen von Gegenständen nennt Wittgenstein „Sachverhalte“ (vgl. TLP 1922: § 2.01), beziehungsweise „Tatsachen“ (vgl. TLP 1922: § 2). Da Tatsachen also die Verbindungen von Gegenständen sind, handelt es sich bei ihnen um Komplexe. Wittgenstein zufolge können wir uns „Bilder“ dieser Tatsachen machen (vgl. TLP 1922: § 2.1), in denen die Elemente im Bild den Gegenständen der Tatsachen korrespondieren (vgl. TLP 1922: § 2.13; § 2.131). Sofern nun die Bilder mit der Wirklichkeit, d.h. Tatsachen, übereinstimmen sind sie wahr. Stimmen sie nicht mit ihnen überein, sind sie falsch (vgl. TLP 1922: § 2.222). Dabei versteht Wittgenstein unter einem Bild

einer Tatsache einen „Gedanken“ (vgl. TLP 1922: § 3). Gedanken können nun in Sätzen sinnlich wahrnehmbar ausgedrückt werden (vgl. TLP 1922: § 3.1). In Sätzen stehen Satzzeichen für die Gegenstände der Tatsachen, auf die sich die Sätze letztlich beziehen (vgl. TLP 1922: § 3.2). Hier merkt Wittgenstein an, dass es sich bei diesen Satzzeichen, um Namen handelt (vgl. TLP 1922: § 3.202; § 3.203). Er macht darauf aufmerksam, dass Namen allein keinen „Sinn“ ausdrücken können (TLP 1922: § 3.142). Das bedeutet Wittgensteins zufolge, dass sie nicht wahr oder falsch sein können. Nur ihre Verbindungen in Sätzen können wahrheitsfähig sein. Auch Wittgenstein hält die Namen, sofern sie sich auf einfache Gegenstände beziehen, für nicht weiter analysierbar. Sie sind das Ende der Analyse (vgl. TLP 1922: § 3.26; § 3.261). Daher nennt Wittgenstein auch einen Satz, der nur aus der Verbindung von „einfachen Zeichen“, d.h. Namen, besteht „vollständig analysiert“ (vgl. TLP 1922: § 3.201). Die vollständig analysierten Sätze werden von Wittgenstein als Elementarsätze bezeichnet (vgl. TLP 2010: § 4.22).

Die Urbestandteile des *Theaitetos* entsprechen also den Gegenständen des *Tractatus*. Sie werden als einfache Elemente vorgestellt, die sich nicht weiter zergliedern lassen und mit einfache Zeichen benannt werden. Die Namen, die sich auf Urbestandteile und Gegenstände beziehen, sind selbst nicht wahrheitsfähig. Erst durch deren Verbindung lassen sich wahrheitsfähige Sätze formulieren.

4. Argument für den logischen Atomismus

Welche Gründe sprechen nun für den logischen Atomismus? Warum muss man bei der Analyse wahrheitsfähiger Sätze schließlich auf Namen stoßen, die sich nicht weiter analysieren lassen? Warum sollte die Analyse nicht unendlich weiterführbar sein?

Wittgenstein führt an, dass sonst die Wahrheit oder Falschheit von Sätzen unbestimmt wäre:

- 2.0211 Hätte die Welt keine Substanz, so würde, ob ein Satz Sinn hat, davon abhängen, ob ein anderer Satz wahr ist.
- 2.0212 Es wäre dann unmöglich, ein Bild der Welt (wahr oder falsch) zu entwerfen. (TLP 1922)

Mit Substanz sind hier die Gegenstände der Elementarsätze gemeint (vgl. TLP 1922: § 2.021). Wittgenstein geht somit davon aus, dass die Wahrheit aller Sätze von immer grundlegenden Sätzen abhängen würde, wenn es keine Elementarsätze gäbe, die nur aus der Verknüpfung von Namen bestehen, welche auf einfache Gegenstände Bezug nehmen, die man letztlich bei der Analyse von wahrheitsfähigen Sätzen erreicht. Dies liegt daran, dass Wittgenstein der Auffassung ist, dass die Wahrheit eines Satzes aus der Wahrheit eines anderen folgt, wenn ersterer dieselben Wahrheitsgründe aufweist (vgl. TLP 1922: § 5.121). Wittgenstein nimmt deshalb an, dass es Sätze geben muss die ihre eigenen Wahrheitsgründe aufweisen. Bei diesen Sätzen handelt es sich um Elementarsätze. Gäbe es diese Sätze nicht, könnte nicht mehr sichergestellt werden, dass mit unseren beschreibenden Sätzen überhaupt etwas Wahres formuliert werden könnte. Jemand der an der Wahrheit aller Sätze zweifelt, wäre immer auf die Wahrheitsgründe weiterer, fundamentaler Sätze verwiesen und so *ad infinitum*. Die Elementarsätze fungieren somit als Regressstopper. Der Zweifler kann auf die Elementarsätze verwiesen werden, die ihre eigenen Wahrheitsgründe aufweisen, da sie in direktem Kontakt mit den Gegenständen stehen.

5. Platons Argument gegen den Logischen Atomismus

Damit hat Wittgenstein ein Argument für den logischen Atomismus in der Hand. Doch wie steht es um die Plausibilität der Existenz logischer Atome? Wie Gabriel (2012: 55) bemerkt, ist eine der Pointen des *Theaitetos*, dass es keine logischen Atome gibt. Wie oben ausgeführt, ist eine wesentliche Eigenschaft der logischen Atome ihre Unerkennbarkeit; da die Namen, die die logischen Atome benennen, nicht wahrheitsfähig sind, können sie auch nicht erkannt werden. Genau auf diesen Umstand zielt Sokrates Kritik im *Theaitetos* ab (McDowell 1973: 240).

Sokrates zweifelt an der These, dass die logischen Atome unerkennbar sein sollen, aber ihre Verknüpfungen erkennbar (vgl. Tht. 1992: 202a). Sokrates versucht am Beispiel einer Silbe und ihrer Buchstaben ein Dilemma zu konstruieren, das gegen die Plausibilität des logischen Atomismus spricht. Das Dilemma besagt, dass eine Silbe und ihre Buchstaben entweder beide erkennbar oder beide unerkennbar sein müssten (vgl. McDowell 1973: 241).

Um das Dilemma zu begründen, versucht Sokrates zu zeigen, dass ein Ganzes identisch mit seinen Teilen sein muss (vgl. Th. 1992: 204b-205a). Unter dieser Voraussetzung, ergibt sich, dass wenn eine Silbe erkennbar ist auch ihre Urbestandteile, d.h. ihre Buchstaben, erkennbar sein müssen. Dies erklärt sich dadurch, dass im Fall der Identität von Teilen und Ganzem alle Prädikate, die auf das Ganze zutreffen, auch auf seine Teile zutreffen müssen (vgl. McDowell 1973: 242). Möchte man dieses Resultat vermeiden, ist man zur Aufgabe der Annahme gezwungen, dass die Teile einer Silbe Buchstaben sind. Man müsste also bestreiten, dass eine Silbe Teile aufweist (vgl. McDowell 1973: 243). Somit müsste aber eine Silbe selbst als Urbestandteil gelten und dadurch unerkennbar sein. Folglich wären Silbe und Buchstaben unerkennbar (vgl. McDowell 1973: 246).

Überträgt man diese Überlegungen auf die Urbestandteile und Erklärungen, um die es im Traum geht, den Sokrates referiert, ergibt sich auch hier das Dilemma. Wenn Erklärungen aus Urbestandteilen zusammengesetzt sind, dann müssen alle Prädikate, die für die Erklärungen gelten auch, für die Urbestandteile gelten. Da Erklärungen erkennbar sind, müssen auch ihre Urbestandteile erkennbar sein. Um diese Konsequenz zu vermeiden, müsste man behaupten, Erklärungen seien nicht aus Teilen zusammengesetzt, aber dann würde es sich bei ihnen selbst um Urbestandteile handeln. Damit wären aber auch Erklärungen unerkennbar. Beide Seiten des Dilemmas widersprechen also der Annahme des Traumes.

In Anbetracht des Gegenarguments lässt sich für Wittgensteins logischen Atomismus nicht mehr behaupten, dass Sätze wahrheitsfähig sind, aber die Namen, aus denen Sätze zusammengesetzt sind, nicht. Wenn Sätze wahrheitsfähig und identisch mit ihren Teilen sind, dann müssen auch ihre Bestandteile wahrheitsfähig sein. Die Annahme von logischen Atomen scheint also ein inkonsistentes Postulat zu sein.

Wie McDowell (1973: 250-251) jedoch bemerkt, beruht das Dilemma auf der Annahme, dass jedes Ganzes identisch mit seinen Teilen ist. Diese Annahme lässt sich jedoch zurückweisen. Nicht jedes Ganze kann identisch mit seinen Teilen sein, da es sich bei einigen Ganzen um Teile in einer bestimmten Ordnung handelt. So unterscheiden sich die Silben /mur/ und /rum/ in der Ordnung ihrer Teile. Sie können somit nicht bloß aus ihren Teilen bestehen.

Wittgenstein ist sich dieses Punkts im *Tractatus* bereits bewusst. Er bemerkt, dass die „Konfiguration“ der Gegenstände entscheidend für den Sachverhalt ist (vgl. TLP 1922: § 2.072), und dass die „Art und Weise“ der Anordnung der Gegenstände, die „Struktur“ des Sachverhalts ausmachen (vgl. TLP 1922: § 2.032). Da nun in Sätzen die Ordnung der Namen die Ordnung der Gegenstände in Sachverhalten widerspiegelt, ist auch die Ordnung der Namen entscheidend für einen Satz. Ein Satz besteht also nicht nur aus seinen Namen, sondern eben auch aus der Anordnung der in ihm verwendeten Namen. Hier unterscheiden sich also der logische Atomismus des *Tractatus* und der des *Theaitetos* (vgl. McDowell 1973: 234).

6. Wittgensteins Revision des logischen Atomismus

Trotzdem verneint Wittgenstein im *Tractatus* nicht explizit, dass jedes Ganze identisch mit seinen Teilen sein muss. In den *Philosophischen Untersuchungen* hingegen fragt er ausdrücklich, ob nicht etwas bei der Zergliederung von Sätzen, d.h. ihrer Analyse in Namen, verloren geht (vgl. PU 2022: § 63). Darüber hinaus fragt er: „Ja, der Besen wird zerlegt, wenn man Stiel und Bürste trennt; aber besteht darum auch der Befehl, den Besen zu bringen, aus entsprechenden Teilen?“ (PU 2022: § 60). Damit zweifelt er an der im *Tractatus* getroffenen Annahme, dass die Ordnung der in Sätzen verwendeten Namen die Ordnung der Gegenstände in Sachverhalten widerspiegelt. Nur weil sich ein Besen in Teile zerlegen lässt, heißt das noch nicht, dass sich auch ein Teil eines Satzes in der gleichen Art und Weise zerlegen lässt.

Damit korrespondieren die Struktur eines Satzes und die eines Sachverhalts nicht mehr in jedem Fall miteinander. Die Korrespondenz von Satzordnung und der Ordnung des Sachverhalts ist aber der Grund, warum sich dem *Tractatus* zufolge Sätze nicht ohne weiteres auf ihre Teile reduzieren lassen. Schließlich spiegelt die Satzordnung die Ordnung der Sachverhalte. Diese würde bei einer Reduktion auf ihre Teile verloren gehen. Wenn nun aber Wittgenstein in den *Philosophischen Untersuchungen* behauptet, dass sich Satzordnung und Ordnung des Sachverhalts ohnehin nicht notwendig entsprechen, dann muss es einen anderen Grund geben, warum sich ein Satz

nicht auf seine Teile reduzieren lässt. Wittgenstein verweist zunächst nur darauf, dass es sich um ein anderes Sprachspiel handelt, wenn ein Satz in seine Teile zergliedert wird:

In wiefern sind die Zeichen dieses Sprachspiels einer Analyse bedürftig? Ja, in wie wieweit *kann* das Spiel durch (48) ersetzt werden? – Es ist eben ein *anderes* Sprachspiel; wenn auch mit (48) verwandt. (PU 2022: § 64)

Es kann also festgehalten werden, dass Wittgenstein nicht mehr davon ausgeht, dass sich Sätze immer weiter analysieren lassen, bis man auf die ihnen zugrunde liegenden Sätze stößt, die nur noch aus Namen bestehen. Sätze lassen sich eben nicht in jedem Fall weiter analysieren, da etwas bei der Analyse verloren gehen kann, d.h., das Sprachspiel kann sich verändern.

Auch revidiert er seine Meinung darüber, dass „Namen eigentlich das Einfache bezeichnen“ (PU 2022: § 41). Wittgenstein problematisiert in den *Philosophischen Untersuchungen*, dass die Aussage, etwas sei „einfach“ oder „zusammengesetzt“, zu unbestimmt ist. Es müsse zunächst geklärt werden, um welche Art der Zusammensetzung es sich handelt (vgl. PU 2022: § 47). Man könnte sich schließlich ein rotes Quadrat vorstellen und fragen, ob es ein einfaches Element oder aus weiteren Elementen zusammengesetzt ist. Wenn man nun unter weiteren Elementen weitere Quadrate versteht, dann ist es ein einfaches Element, denn es ist nur *ein* Quadrat. Versteht man aber unter weiteren Elementen Formen und Farben, dann ist es aus der Form eines Quadrates und der Farbe Rot zusammengesetzt (vgl. PU 2022: § 48).

Außerdem schränkt er die Behauptung des *Tractatus* ein, einfache Elemente ließen sich nicht erklären, sondern nur benennen. In den *Philosophischen Untersuchungen* wendet er ein, dass es auf die Situation ankommt, ob ein Element nur benannt werden kann. Wenn ein Name benutzt wird, um etwas zu beschreiben, besitzt das Element, das zu Beschreibung benutzt wird, selbst zwar keine Beschreibung, aber das bedeutet nicht, dass es niemals beschrieben werden kann. Das Element könnte in einer anderen Situation zum Gegenstand einer Beschreibung werden. Deshalb schreibt Wittgenstein:

darum wäre es hier seltsam zu sagen, das Element könne man *nur* benennen! Benennen und Beschreiben stehen ja nicht auf *einer* Ebene: Das Benennen ist eine Vorbereitung zur Beschreibung. (PU 2022: § 49)

Dass Benennen und Beschreiben nicht auf einer Ebene liegen bedeutet, dass die Elemente, die auf der einen Ebene beschrieben werden, Elemente auf einer anderen Ebene benötigen, die sie beschreiben. Die Elemente dieser anderen Ebene werden in diesem Fall nur benannt, aber nicht beschrieben. Dennoch können sie beschrieben werden, indem Elemente einer weiteren Ebene herangezogen werden.

Deshalb sagt Wittgenstein in Anlehnung an den *Theaitetos*, dass man den einfachen Elementen, die auf der einen Ebene nur benannt werden können, „weder Sein noch Nichtsein beilegen“ (PU § 50) könne. Einfache Elemente werden genutzt, um überhaupt Aussagen über Elemente einer anderen Ebene zu treffen. Somit werden sie auch gebraucht, um Seinsaussagen zu treffen. Also sind sie die Bedingung, um Seinsaussagen treffen zu können und sie selbst können nicht gebraucht werden, um ihnen Sein zuzuschreiben. Um ihnen Sein zuzuschreiben, werden wieder Elemente einer anderen Ebene benötigt. Wittgenstein vergleicht die einfachen Elemente, die als Mittel gebraucht werden, um Aussagen zu treffen, mit dem Urmeter in Paris, das dazu genutzt wird, um die Länge eines Meters festzulegen. Da das Urmeter in Paris die Bedingung ist, um die Länge eines Meters feststellen zu können, kann von ihm selbst nicht gesagt werden, dass es ein Meter lang sei (vgl. PU 2022: § 50). Somit entfernt sich Wittgenstein in den *Philosophischen Untersuchungen* also von Annahmen des logischen Atomismus.

7. Fazit

Der im *Theaitetos* als Traum bezeichneten Theorie und der Position Wittgensteins im *Tractatus* ist die Annahme logischer Atome gemeinsam, die sich nicht weiter analysieren lassen und durch einfache Zeichen benannt werden. Aus diesen einfachen Zeichen setzen sich schließlich wahrheitsfähige Sätze zusammen. Für die Plausibilität dieser Annahme spricht der Umstand, dass logische Atome dafür sorgen, dass einige Sätze ihre eigenen Wahrheitsgründe aufweisen. Sätze, die nur aus einfachen Namen bestehen, stehen in direktem Kontakt mit logischen Atomen. Da Namen allein nicht wahrheitsfähig sind, hängt die Wahrheit von Sätzen, die nur aus diesen bestehen, von keinen anderen Sätzen ab. Damit weisen sie ihre eigenen Wahrheitsgründe auf. Nichtsdestotrotz konnte durch die Kritik des Sokrates

im Theaitetos gezeigt werden, dass die Annahme von logischen Atomen, die durch einfache Zeichen benannt werden, nur unter der Voraussetzung plausibel ist, dass nicht jedes Ganze identisch mit seinen Teilen ist. In dieser Hinsicht scheint jedoch Wittgensteins Position im *Tractatus* nicht betroffen zu sein. Der Umstand, dass Wittgenstein dafür hält, dass ein Satz aus Namen und einer gewissen Ordnung besteht, zeigt, dass hier nicht jedes Ganze auf seine Teile reduziert werden kann.

In den *Philosophischen Untersuchungen* kritisiert Wittgenstein jedoch die Annahmen des logischen Atomismus. Erstens behauptet Wittgenstein nicht länger, dass sich Sätze auf immer fundamentalerer Sätze reduzieren lassen, sondern er ist der Auffassung, dass etwas bei ihrer Analyse verloren geht. Zweitens ist er auch nicht mehr der Auffassung, dass es einfache Gegenstände oder einfache Elemente in einem absoluten Sinn gibt. Ob ein Gegenstand einfach oder komplex ist, hängt dem späten Wittgenstein zufolge von bestimmten Kontextbedingungen ab. Damit revidiert Wittgenstein seine frühere Position.

References

Gabriel, Markus (2012) *Die Erkenntnis der Welt. Eine Einführung in die Erkenntnistheorie*, 4. Auflage, Freiburg in Breisgau: Verlag Karl Alber.

Plato (1973) *Theaitetus*, Hrsg. von John McDowell, übersetzt mit Kommentar von John McDowell, Oxford: Oxford University Press.

Platon (1992) *Theaitetos*, in: ders.: Sämtliche Werke Bd. 3, Hrsg. von Ursula Wolf, Übersetzung von Friedrich Schleiermacher, Leipzig: Rohwolt.

Proops, Ian (2017) "Wittgenstein's Logical Atomism", The Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/archives/win2017/en/ries/wittgenstein-atomism/> (letzter Abruf: 05.04.2022)

Russel, Bertrand (2010) *The Philosophy of Logical Atomism*, Oxon: Routledge Classics.

Ryle, Gilbert (1990) "Logical Atomism in Plato's 'Theaitetus'", *Phronesis* 35.21-46.

Wittgenstein, Ludwig (1922) *Tractatus Logico-Philosophicus*, London: Kegan Paul, Trench, Trubner & Co.

Wittgenstein, Ludwig (2022) *Philosophische Untersuchungen*, Suhrkamp: Frankfurt am Main.

Overcoming Modal Skepticism via Conceptual Engineering

Krzysztof Sękowski (Warsaw, Poland)

Abstract

The paper defends the view that that a significant part of metaphysics should be understood as conceptual engineering, and explores its epistemological advantages and metaphysical implications. It discusses the challenge raised by moderate modal skepticism, which points out the lack of reliable methods for verifying modal statements through thought experiments. As I argue, a normative stance on metaphysical methodology, understanding it as engaging in conceptual engineering project, justifies the use of this method. By adopting an approach where the method of cases offers reasons for adopting specific normative constraints on concept usage, the method becomes justifiable. Consequently, it can be employed to justify the revision of conceptual schemes and, in turn, validate certain modal claims about entities falling under the scrutinized concepts. The paper explores how this approach can methodologically enrich conceptual engineering-focused perspectives in metaphysics and examines its metaphysical consequences. In particular, it demonstrates that despite the mind-dependency of the method of cases results, it still leaves room for a metaphysically realistic viewpoint.

Metaphysics is concerned to a large extent with questions about modal facts. However, its methodology has been challenged by modal skepticism (e.g. van Inwagen 1998; Nozick 2001; Machery 2017, 2023), which casts doubt on our ability to justify claims about necessities, fundamentally questioning the extent of our metaphysical understanding. Therefore, modal skepticism not only impacts the methods of evaluation of metaphysical theories but also calls into question the value of existing theories concerning the structure of reality itself, challenging the foundation upon which metaphysical knowledge is built.

The recent version of modal skepticism (Alexander, Weinberg 2007; Machery 2017, 2023) highlights the foundational role of thought experiments in exploring metaphysical modal facts. It challenges the assumption that thought experiments can reliably lead to knowledge about metaphysical possibilities and necessities due to demographic differences and framing effects regarding verdicts on hypothetical cases.

In the paper, I discuss a reinterpretation of metaphysical methodology, arguing for a perspective where debates on metaphysical modality are understood as being partly normative and ameliorative. Under this understanding, the proper way of doing metaphysics is engaging in a conceptual engineering enterprise. I defend this approach by discussing a re-

interpretation of thought experiments that is coherent with it. I show how adopting a normative account of the methodology of metaphysics can justify and show the method of thought experiments in a different light. According to this view, when we argue about the essential attributes of entities, we are not merely describing how things are but advocating for how we should conceive of them. In particular, by considering hypothetical examples while evaluating metaphysical theories, we impose normative constraints on our concepts. As I demonstrate, the arguments made by moderate modal skeptics require the adoption metaphysical methodology seeing it as conceptual engineering.

Section 1 presents the arguments of modal skeptics regarding the reliability of the method of cases. Section 2 introduces a motivation for reinterpretation of metaphysical methodology, suggesting that modal disputes concern normative claims about how we should construct conceptual schemes that partly determine modal properties of entities rather than merely descriptive facts about what these properties are. Finally, section 3 presents how the use of counterexamples in thought experiments can be justified within the proposed reinterpreted metaphysical approach and discusses its metaphysical consequences, especially concerning the mind-(in)dependence of its results.

1. The Challenge from Moderate Modal Skepticism

The starting point for modal skepticism is the observation that philosophers, when arguing in favor of a claim concerning metaphysical possibility or necessity, often refer to hypothetical situations and judge whether a certain state of affairs is possible, impossible, or actual. For example, to argue that placing a stone on Mars is possible, one might consider a scenario where a stone is placed on Mars and conclude that it is plausible (probably yes). Similarly, to argue that a bachelor being married is impossible, one could imagine a bachelor getting married and then assess whether he remains a bachelor (likely not, setting aside legal technicalities or undisclosed divorces).

The thought experiment is a method in which we examine a hypothetical scenario and make judgments about it for certain epistemic purposes. Such purposes vary; some thought experiments illustrate theories, others provide puzzles that advance scientific or philosophical discussions. A particularly significant use of thought experiments, the method of cases, is to provide

counterexamples (think e.g. about Gettier Cases, or thought experiments discussed within the discussion on personal identity). Below, when I will talk about thought experiments I'll have that particular kind in mind, limiting my considerations to that one, although influential, kind of thought experiments.

As a matter of fact, philosophers use the method of cases to establish modal facts. For instance, when Chalmers (1997) argues against physicalism's claim that consciousness is necessarily reducible to physical states, he presents a counterexample through a thought experiment involving an entity with the same physical structure as a conscious person but lacking phenomenal experience. Similarly, Gettier (1963) challenges the notion that knowledge is necessarily justified true belief by offering thought experiments where someone has a justified true belief on some proposition, but does not know it. These examples highlight the initial step in the argument for modal skepticism: metaphysics relies heavily on thought experiments to justify modal claims.

However, some proponents of modal skepticism argue that this method is unreliable. They point to empirical studies showing demographic differences or framing effects on thought experiment verdicts (Machery 2017; Feltz & Cokely 2019; Machery & Stich 2023; Sękowski et al. 2023; but see also: Knobe 2021; 2023). They further argue that if we reject the idea that e.g. the personality of the thought experimenter determines whether indeterminism is necessary for free will, we should conclude that investigating hypothetical scenarios is an unreliable tool for establishing modal truths. Therefore, just as we would discard measuring instruments in scientific practice if they were imprecise or inadequate, so too should we abandon the method of cases. This brings modal skeptics to the following conclusion: the methodology of metaphysics is in a serious trouble, since the foundational method to establish modal truths should be rejected (Alexander, Weinberg 2007; Machery 2017, 2023). An important feature of this critique I aim to overcome is the presupposition inherited from the traditional philosophical methodology picture that the primary aim of method of cases is to uncover modal facts, similarly to scientific experiments that seek to discover empirical facts.

Note that the skeptical challenge refers specifically to the method of cases, and a specific kind of modality, namely, metaphysical modality (as opposed to e.g. nomological modality). This specificity is why Machery labels such a

perspective moderate modal skepticism. Nevertheless, given that metaphysical modality is central to the interests of metaphysics, just as thought experiments are fundamental to its methodology, the argument presented poses a significant challenge to both metaphysics and its methodology.

2. From Challenge to Change: Towards Normative Metaphysics

The solution to addressing modal skepticism is to reinterpret metaphysical methodology. In this section, I explore the motivations behind the view that metaphysics should be viewed as conceptual engineering. I discuss Machery's proposal, suggesting an alternative interpretation of thought experiments, and I show why his approach, while enriching the project of normative metaphysics, falls short for those seeking to employ the method of cases to justify certain modal claims.

Conceptual engineering is a methodological approach that has gained a lot of attention in the last few years. According to it, philosophers rather than being focused on establishing conceptual truths, should aim at evaluating or changing our concepts understood in a linguistic or psychological way (Isaac et al. 2022). Thereby, it provides a way of thinking about philosophical problems in a normative way rather than a descriptive one.

Such an approach has been adopted in the past by a lot of philosophers, among others by Carnap in his project of explication (Carnap 1950), the proponents of the Lvov-Warsaw School, while providing arguments for constructing concepts (Łukasiewicz 1905/2022) or Foucault in his projects of conceptual genealogy (Foucault 1971). Additionally, it has been argued that actually a lot of philosophical debates might be understood as metalinguistic negotiations (Plunkett 2015, Thomasson 2017). When it comes to the advantages of such an approach, a normative view of philosophical methodology, seeing it as a kind of conceptual engineering, has been supported by a variety of its methodological, epistemological, and metaphysical benefits (Thomasson 2017; 2020, forthcoming; Sękowski 2022a). Let us focus now on a particular epistemological advantage of such an approach: a way to address the challenge of modest modal skepticism.

A step towards such a methodological turn has been made by Machery himself, (2017), who also defended the recent version of modest modal skepticism. He states that since we can't defend the method of cases as providing justification to modal claims, we should think about whether we can use that method to justify a prescriptive project that aims at replacing concept-constitutive beliefs with new ones in line with a certain (social, ethical, theoretical, etc.) aims. The method of cases 2.0, as he calls it, as an empirical method of studying responses to thought experiments can provide a first step for this project. It reveals possible ways in which the concept may play the role, by showing what kind of inferences people are inclined to draw from the use of that concept. That step might fuel the conceptual engineering either by suggesting desired revisions, if the discovered inferences are in line with certain aims, or revealing the needed-to-be-fixed fallacies, if, given certain aims, these inferences are intrusive.

Machery's method of cases 2.0 is undeniably a valuable tool for a proponent of the normatively-oriented metaphysics. However, it also significantly diverges from the goals of the original method of cases. The difference between the method of cases and its 2.0 version lies not solely in the shift from revealing conceptual or modal facts to becoming part of an ameliorative project. The difference lies in the fact that while the method of cases 2.0 can provide reasons to initiate concept revision or inspire it, it doesn't directly justify any modal claims traditionally the method of cases was aimed at. Furthermore, while it is an intriguing approach for the future, it does not apply to past philosophical practices, since as an empirical method, and contrary to the way in which thought experiments have been used, the method of cases 2.0. requires a huge sample of verdicts on thought experiments.

Presenting a method that not only enriches the methodological toolkit of conceptual engineering-focused metaphysics but also captures the core of previous philosophical ways of acting lends greater plausibility to the project of understanding metaphysics through the lens of conceptual engineering. This insight aligns with arguments posited by those who argued that conceiving philosophical methodology as conceptual engineering is not only a fitting guide for future endeavors but also provides a plausible interpretation of what philosophers have largely been engaged in history (Thomasson 2017, forthcoming; Andow 2020; Sękowski 2022a, 2022b). The significant advantage

of this perspective is that it introduces a new picture of philosophical methodology that remains compatible with the history of philosophy, treating conceptual engineering as a seamless continuation of the philosophical tradition. Consequently, conceptual engineering can draw upon the richness of past philosophical achievements, enhancing its methodological foundation. Interpreting past theories and arguments along the conceptual engineering lines has the potential to inform future projects by offering results received in the past and argumentative strategies applicable to forthcoming philosophical enterprises.

3. Method of Cases in Service of Normative Metaphysics

Similarly to Machery, I propose reinterpreting the method of cases as a part of conceptual engineering endeavour. In contrast to him, I aim to show how this method can be reinterpreted to justify its past uses, thereby connecting the conceptual engineering turn more closely to the historical philosophical tradition. The defended interpretation illustrates how the method of cases can justify conclusions in metaphysical arguments. Let's turn to the details of this approach, its ability to address skeptical challenges, and its implications for metaphysics.

Contrary to the presupposition of the debate on the method of cases, according to my proposal, the primary aim of the method of cases is not to uncover modal facts but to offer reasons for reevaluating our conceptual frameworks. These changes concerning the modal features of entities under ameliorated concepts result in modifications to the modal discourse, and the way in which these concepts are used. Normative arguments are presented for adopting a specific conceptual scheme that entails certain modal commitments. These arguments don't rely on direct insight into modal reality but are justified by expectations of the roles scrutinized concepts or conceptual schemes should play, advocating for a shift in the norms governing their use (see Sękowski 2022a; 2022b).

To grasp this idea, note that the textual analysis of a lot of thought experiments reveals that when presented, their verdicts aren't solely justified by their obviousness but by arguments. Proponents of the so-called mischaracterization objection argue that intuitions don't play a justificatory

role in the method of cases or in philosophical methodology overall (see, for example, Cappelen 2012; Deutsch 2015; Horvath 2022; 2023). Since intuitive verdicts on thought experiments are justified by arguments independent of their intuitiveness, there's no need to rely on intuition as a source of justification for the claim a thought experiment aims to support. According to proponents of the mischaracterization objection, this addresses the skeptical challenge. If thought experiment verdicts don't justify the modal claims concluded from arguments with their help, empirical results on the unreliability of those verdicts don't undermine the use of the thought experiment method.

However this response comes into trouble when considering the justification for the premises of arguments supporting thought experiments' verdicts. Consider the argument that Smith in Gettier's thought experiments doesn't know that *p* because he is right about *p* merely by luck (Horvath 2022). As e.g. Chudnoff (2017) or Nado (2016) suggest, the premise that one cannot be right on *p* by sheer luck if one knows it seems to be justified by its intuitiveness, not by further justification. In response, Horvath argued that there's no experimental philosophy works on intuitions about general features of concepts, so although we can be skeptical towards the reliability of our verdicts about thought experiments, there's no reason to doubt the reliability of our intuitions of concept's general features (see Horvath 2023). However, this answer is unsatisfactory due to research beyond experimental philosophy, particularly from developmental psychology, that show significant demographic variations in judgments about the general features of concepts, e.g., knowledge (Baxter Magolda 200; Karabenick, Moosa 2005).

Despite the problems with the mischaracterization objection, we can adopt its perspective that verdicts do not serve as evidence for thought experiments' conclusions. We can also shift our focus to arguments that justify these verdicts, as they argue. However, to effectively address the skeptical challenge, we should move away from looking for a reliable source of evidence in the method of cases and instead adopt a view that demonstrates why it's useful to embrace a particular view on a scrutinized concept. I call for discarding the reliability-centric approach to thought experiment justification. The reason for this is, however, not the belief that our intuitions about thought experiment verdicts constitute modal facts (for such accounts, see Thomasson 2012; Ásta

2013). Even if we don't embrace them, adopting the stance that arguments formulated within the method of cases provides reasons to revise a conceptual scheme based on our needs makes the tracking of modal facts unnecessary. Consequently, there's no need to require reliability for thought experiment verdicts. The justification for the use of the method of cases lies in our expectations regarding the concept under scrutiny or the broader functions the entire conceptual scheme aims to fulfill. If one shares or finds these expectations useful, the entire argument is justified, pushing the project of ameliorating a certain concept forward.

This interpretation of the method of cases aligns with approaches to conceptual engineering that uses the functions a concept should serve as a guide for the project (e.g. Plunkett, Sundell 2013; Simion, Kelp 2020; Thomasson 2020; Queloz 2021; Nado 2021). Besides the methods outlined in such accounts, it illustrates how a traditional and widely-used method, like the method of cases, can be incorporated into these projects, adding to the toolkit of metaphysicians willing to think about that discipline along the lines of conceptual engineering that is function-based.

Let us focus on the metaphysical consequences of this perspective. A key and broad consequence is that claims about modal reality end up being partly dependent on our needs. For instance, whether it's necessary for a "conscious agent" to be a biological organism might be partly tied to what the concept of a "conscious agent" is meant to achieve for us. The purposes steering our current interest in the concept can vary widely, ranging from legal or scientific concerns to ethical or entertainment-related ones (see Burgess, Plunkett 2013). Consequently, the answer to the question of the necessary properties of consciousness may differ for those approaching the concept for legal purposes, considering the challenges of AI development, those developing software for entertaining online games, or cultural anthropologists interpreting cultures that worship personal God, or animated members of flora.

However, at the same time, it's crucial to recognize that despite metaphysical claims depending on contingent purposes, once we establish our expectations for the scrutinized concept or entire conceptual framework, it becomes independent of our choices whether a specific concept revision succeeds in fulfilling the established functions. Therefore, while, for instance, the decision

to focus on the concept of consciousness for legal purposes, defending the rights of a certain group of artists in response to AI-generated products, is a choice, it's not a choice whether the law incorporating that concept successfully ensures income for artists when someone uses an AI-generated product based on someone's work (for similar arguments with respect to conceptual engineering in general see: Thomasson (2020)).

The normative aspect of conceptual engineering-focused metaphysics is expressed through our expectations for the conceptual scheme. However, the application of this conceptual scheme to reality and its success in fulfilling its role are mind-independent. Moreover, the fact that the justification within the method of cases is based on normative reasons doesn't rule out that other methods in metaphysics might refer to a kind of justifications that refer e.g. to claims about the nature of things or relations between them. In this regard, the proposed project is realistic in a crucial sense and aligns with various realistic approaches which accept to some extent the mind-dependency with respect to the choice of conceptual scheme or the way in which we're interested in the world (e.g., Putnam 1987, Jago 2023). Nonetheless, it adopts a more epistemologically responsible approach, as it avoids an ambition to justify modal conclusion of the method of cases solely with reference to mind-independent facts about modal reality, while providing reasons to accept certain modal claims. This is just a more self-aware way to fulfill our needs of modal inquiry with the help of the method of cases.

References

Alexander, Joshua; Weinberg, Jonathan (2007) "Analytic epistemology and experimental philosophy", *Philosophy Compass*, 2(1), 56–80.

Ásta (2013) "Knowledge of essence: The conferralist story", *Philosophical Studies*, 166, 21-32.

Baxter Magolda, Marcia B. (2001) *Making their own way: Narratives for transforming higher education to promote self-development* (1st ed.). Stylus.

Burgess, Alexis; Plunkett, David (2013) "Conceptual ethics II", *Philosophy Compass*, 8(12), 1102–1110.

Carnap, Rudolf (1950) "On explication", in: *Logical foundations of probability*. University of Chicago Press.

Chalmers, David (1997) *The conscious mind: In search of a fundamental theory*, Oxford Paperbacks.

Chudnoff, Elijah (2017) "The reality of the intuitive", *Inquiry*, 60(4), 371-385.

Feltz, Adam; Cokely, Edward (2019) "Extraversion and compatibilist intuitions: A ten-year retrospective and meta-analyses", *Philosophical Psychology*, 32(3), 388-403.

Foucault, Michel (1971) "Nietzsche, la généalogie, l'histoire. In S. Bachelard", in: *Hommage à Jean Hyppolite* (pp. 145–172). Presses Universitaires de France.

Gettier, Edmund (1963) "Is knowledge justified true belief?", *Analysis*, 23(6), 121-123.

Horvath, Joachim (2022) "Mischaracterization reconsidered", *Inquiry*, 1-40.

Horvath, Joachim (2023) "Intuitions in experimental philosophy", in: *The compact compendium of experimental philosophy*.

Isaac, Manuel Gustavo; Koch, Steffen; Nefdt, Ryan (2022) "Conceptual engineering: A road map to practice", *Philosophy Compass*, 17(10), e12879.

Jago, Mark (2023) "Knowing what it is", *Philosophical Studies*, 1-16.

Karabenick, Stuart; Moosa, Samira (2005) "Culture and personal epistemology: U.S. and Middle Eastern students' beliefs about scientific knowledge and knowing", *Social Psychology of Education*, 8(4), 375–393.

Knobe, Joshua (2021) "Philosophical intuitions are surprisingly stable across both demographic groups and situations", *Filozofia Nauki*, 29(2 (114)), 11-76.

Knobe, Joshua (2023) "Difference and robustness in the patterns of philosophical intuition across demographic groups", *Review of Philosophy and Psychology*, 1-21.

Łukasiewicz, Jan (1905/2022) "Analysis and construction of the concept of cause", in: Jacek Jadacki; Edward Świdorski (Eds.), *The concept of causality in the Lvov-Warsaw school: The legacy of Jan Łukasiewicz* (pp. 3–64). Brill.

Machery, Edouard (2017) *Philosophy within its proper bounds*, Oxford University Press.

Machery, Edouard (2023) "In defense of modest modal skepticism", in: *Epistemology of modality and philosophical methodology* (pp. 359).

Nado, Jennifer (2017) "Demythologizing intuition", *Inquiry*, 60(4), 386-402.

Nado, Jennifer (2021a) "Conceptual engineering, truth, and efficacy", *Synthese*, 198, 1507–1527.

Nozick, Robert (2001) *Invariances*, Harvard University Press.

Putnam, Hilary (1987) *The many faces of realism*, Open Court.

Sękowski, Krzysztof (2022a) "Concept revision, concept application and the role of intuitions in Gettier cases", *Episteme*, 1-19.

Sękowski, Krzysztof (2022b) "Reference the untouchable: On the limits of revising concepts using the method of cases", *Synthese*, 201(1), 4.

Sękowski, Krzysztof; Ziółkowski, Adrian; Tarnowski, Maciej (2023) "Western skeptic vs. Indian realist: Cross-cultural differences in Zebra Case intuitions", *Review of Philosophy and Psychology*, 14(2), 711-733.

Simion, Mona; Kelp, Christoph (2020) "Conceptual innovation, function first", *Noûs*, 54(4), 985–1002.

Stich, Stephen; Machery, Edouard (2023) "Demographic differences in philosophical intuition: A reply to Joshua Knobe", *Review of Philosophy and Psychology*, 14(2), 401-434.

Thomasson, Amie L. (2012) "Experimental philosophy and the methods of ontology", *The Monist*, 95(2), 175-199.

Thomasson, Amie L. (2017) "Metaphysical disputes and metalinguistic negotiation", *Analytic Philosophy*, 58(1), 1-28.

Thomasson, Amie L. (2020) "A pragmatic method for normative conceptual work" In: Alexis Burgess; Herman Cappelen, David Plunkett (Eds.), *Conceptual engineering and conceptual ethics*, Oxford University Press, 435-458.

Thomasson, Amie L. (forthcoming). *Rethinking Metaphysics*, Oxford University Press.

Van Inwagen, Peter (1998) "Modal epistemology", *Philosophical Studies*, 67-84.

Acknowledgement

I would like to thank Amie Thomasson, Twan Stielke and Niklas Dahl for reading previous versions of this paper and for a thoughtful discussion of the material during my research visit at Dartmouth College in 2024. The visit and the work were funded by the Kosciuszko Foundation Grant.

Regularities and Social Practices: Reconsidering Wittgenstein's Rule-Following Argument

Kanit (Mitinunwong) Sirichan (Bangkok, Thailand)

Abstract

Wittgenstein's rule-following argument captures the underlying problem concerning how meaning gets attached to what we do. Besides the debate on rules as normative constraints of meaning in the philosophy of language, the illustration of the problem is also evident in how regularities and necessity play a role in understanding social practice in the philosophy of social science. The contested views on the notion of practice are between a thin view (Humean regularist view) and a thick view (e.g., presumably, Wittgensteinian view). While the thin view is represented as the view that there are mere regularities in behaviour, the thick view takes it that necessities (in the sense of normativities) are inhered in regularities. According to Haslanger (2013), the thickest view can fall into the trap of an intentionally overloaded thick view, namely, the view that there is no gap between the regularities in what we do and the normativity of what we do. Examples of the intentionally overloaded thick view discussed in this paper include the ideas of Winch and Diamond. Based on McDowell's reading of Wittgenstein's rule-following argument, this paper argues for the non-intentionally overloaded thick view where rules are engaged within practices but are objectively characterizable. Arguing for the non-intentionally overloaded thick view, the paper exploits Haslanger's project of doing philosophy, namely, the ameliorative aim. This sort of thick view entails realism about social structure, which may lead to the circularity problem. That is, behaviour regularities are assumed to be the effect of some causal properties of a social structure before it is known whether such properties exist. However, the paper argues that the regularities in question should be taken as phenomenal regularities, which can be stated as counterfactual conditions.

0. Introduction

The fundamental issue of how meaning is attached to our actions or the concept of practice is captured by Wittgenstein's rule-following argument. Among the various types of reading of the argument, McDowell's transcendental reading is the most appropriate for answering the question. In particular, the answer can help illuminate a disputed issue in the philosophy of the social sciences, namely, the role of regularities and necessities in explaining social phenomena. The contested views are between a thin and thick view. As it is generally understood, the thick view holds that normativities or meanings are inherent in regularities in behaviours. The thickest view is what Haslanger (2013:9) refers to as the intentionally loaded view holding that all regularities in behaviours are intentional or laden with meanings. The *thin* view holds that descriptive facts are self-standing, and to make sense of such facts, one needs a value-added schema. The thinnest view

holds that there are mere regularities in behaviours. In contrast to the thin view, the thick view rejects the idea that descriptive facts and their meanings have a detached nature. However, rejecting detachment makes it easy for the thick view to be intentionally overloaded, in which there is no friction between what is right and what seems right. It is precisely what Wittgenstein remarks in PI 2009: §201 as a paradox: there is no distinction between following and conflicting with the rule. A solution can be found in McDowell's interpretation of Wittgenstein's rule-following argument that we require the idea of rules that are critically engaged inside practices or the non-intentionally overloaded thick view (NTV). It is the idea of social practices where the engagement with rules takes shape. Looking at it from the perspective of philosophy's ameliorative aim, the NTV is also more appealing. The ameliorative aim is to address and perhaps rectify social injustice philosophically. As Haslanger says, "... philosophical inquiry is an inquiry into the concepts we (collectively) ought to use." (2013:22).

In this paper, I examine two kinds of thick views, Winch and Diamond, to show how their views are sort of intentionally overloaded thick view and fall into the trap of the rule-following dilemma. I will first briefly introduce McDowell's reading of Wittgenstein's rule-following argument and use it as a gauge for a sufficient account of the NTV. Then, in the second section, I look at the intentionally overloaded thick views, Winch and Diamond, to show that their views cannot avoid the dilemmatic situation. In the last section, I argue that the NTV is most apt for the ameliorative aim of doing philosophy.

1. Wittgenstein's Rule-Following Argument

McDowell's way of reading the rule-following argument provides a good test tool for a sufficient thick account of practice. The point of the rule-following argument (PI 2009: §138-242) concerns the nature and existence of the normative patterns that govern meaning and use. In using words, it seems some normative rules or patterns guide us on the correct and incorrect use. However, when articulating the rule, we face a dilemma. That is to say, if there are rules for the correct use, then this opens to the regress of rule interpretations; conversely, if we deny that there are rules at all, then we are left without a concept of the normativity of meaning, and the distinction

between correct and incorrect disappears. McDowell (1984: 342) calls this situation the dilemma of Scylla and Charybdis. On the one hand, we will face the Scylla if following a rule is governed by a super-rigid rule, the problem of which is the regress of rule-interpretation. On the other hand, if rules do not exist, we are in the Charybdis situation; there will be no normative constraint for following rules. The latter is the situation of “bedrock”, where things can be as they are; nothing seems right or wrong (PI 2009: §217).

The solution to the dilemma lies in an explanation of practice, which is not only an act without interpretation but also rule-governed. Such a solution requires the idea of belonging to “a custom (PI §198), practice (PI § 202), or institution (RFM VI-31)” (1984: 342). An explanation of meaning that is internal to the linguistic community's context is necessary for the account of linguistic use. However, it is the sense of a linguistic community which is “...bound together, not by a match in mere externals (facts accessible to just anyone), but by a capacity for a meeting of minds.” (ibid.:351). The ‘meeting of minds’ is the way one can know another’s meaning without interpretation. So, in that sense, we do not employ the concept of practice or ‘communal practice’ as an interpretation of meaning. Instead, the communal practice itself is the context within which meaning lies.

There are two crucial points here. First, though communal practice is the way out, it does not mean there is no objectivity of meaning. It means that the objectivity of meaning is not autonomous from use within a context. The rejection of the autonomy of meaning is only in the sense that the pattern of use extends itself to new cases without our perspective on the pattern. In other words, a notion of truth-conditions is still *given in the language* we use and understand. Second, the objectivity of meaning involves our epistemic and ontological engagement with an object. Epistemically, the engagement is in our ways of thinking about an object. Ontologically, the object of our direct engagement does exist.

To understand how this is possible, we need to grasp the idea of direct realism. Such an idea is provided by McDowell’s disjunctive account of perceptual experiences. (McDowell 1986:151). The disjunctive account of experiences is the idea that there are no common characteristics between experiences of appearance and experiences of reality. But they are disjunctively interwoven

with each other, namely, “that an appearance that such-and-such is the case can be *either* a mere appearance *or* the fact that such-and-such is the case making itself perceptually manifest to someone.” (McDowell 1982:386). There are perceptual experiences in non-veridical cases (such as delusions, illusions, and hallucinations), but its content is merely non-veridical descriptive content. An important implication is that an independent objective constraint on experiences is possible because of the disjunctive fact, which shows a difference between the two experiences.

McDowell's disjunctive account of experience interprets Wittgenstein as a direct realist. This is different from other ways of realist reading, e.g. Diamond's resolute reading (Diamond 1995, 2012) and contra to other readings like Winch's (1958, 1964). In the next section, I argue how Winch and Diamond's reading fall into the dilemmatic situation of Scylla.

2. The Intentionally Overloaded Thick View

Like McDowell's reading, the realistic spirit that Diamond attributes to Wittgenstein does not reject the objectivity of meaning; instead, it rejects the idea of objectivity independent of our thoughts' characterization. However, such a rejection is at risk of attributing Wittgenstein a relativist view. Diamond's reading tries to steer its way toward non-relativism but unsuccessfully when addressing the issue of incommensurability between conflicting worldviews. This can be seen when she argues against Winch (1964) and Dilman (2002), who take Wittgenstein to be more relativist in nature (see Diamond 2012).

In criticizing Anscombe (1981), Diamond thinks that Wittgenstein's position in dealing with the problem of incommensurability is much more complicated, i.e., criticism from outside different language-games or practices is impossible. For Diamond, there are “rational grounds” for criticizing conflicting worldviews, but not in the sense of “the standards of what counts as rational available independently of and prior to the articulation of thought about conflicting worldviews” (Diamond 2012: 129). Diamond seems to suggest that in criticizing any different or conflicting worldviews, we need to conceptualize their worldview from our own. She uses the practice of “witch-fearing and witch-finding” as an example, presumably in opposition to “witch-denial” in

scientific practice. According to Diamond, judgments about conflicting worldviews, such as those between science and witchcraft, cannot be made "in advance of the conflict itself." (ibid.:128). This means that the justification of a practice needs to be expressed within its unique conceptual framework. Science, as an external practice, is not an adequate justification to support the practice of witchcraft in this case.

The implication is that the conflicting practices are incommensurate, meaning that it is not conceivable to take in criticism from outside each practice. Yet, that sounds far-fetched. The main problem is that the realistic spirit that Diamond ascribes to Wittgenstein is not realistic enough. It is still trapped in the intentionally overloaded thick view, so that changing view requires merely an inferential practice within one's own schema, the problem of which is the loss of an independent normative judgement.

The problem with Diamond's view is similar to Winch's treatment of the rule-following argument. According to Winch's Wittgenstein, understanding meaning is following a rule. This is a grasp of the normativity of meaning. Winch said that "...the notion of following a rule is logically inseparable from the notion of *making a mistake*" and "A mistake is a contravention of what is *established* as correct; as such, it must be *recognizable* as such a contravention." (Winch 1958: 32). The criterion of mistake is established not by "any individual in complete isolation from other individuals.". He says, "For it is contact with other individuals which alone makes possible the external check on one's actions which is inseparable from an established standard." (Ibid.). This means that participation in a community practice, or the internal standard of correctness, clearly defines the idea of normativity of meaning. This point is apparent when Winch discusses Pareto's view on the problem of how, for a sociologist, social phenomena can be studied independently from being participants in such phenomena (Winch 1958: 95). Pareto's answer is to employ a scientific approach to sociology where objective knowledge can be gained. That is possible because of Pareto's naturalistic assumption, that is, social events can be viewed as physical events; therefore, being participants in the social events as the object of study can be ignored. But Winch's criticism is that Pareto's answer is absurd because if one studies the social phenomena the same way as physical phenomena, "the events he is studying lose altogether their character as *social events*" (Winch 1958: 108).

Social events involve participants' way of life. So, Winch says, "It is not open to him arbitrarily to impose his own standards from without." (Ibid.). In parallel with Diamond, the practice of natural sciences is not the rational standard for making judgements on non-scientific practice (Winch 1958:102). However, for Winch, this does not mean that scientific practice in itself is not legitimate. The reason for him might be that philosophy is an "uncommitted enquiry" (Ibid.). In contrast to the philosophy's ameliorative aim, Winch takes philosophy's task as concerning the elucidation of various forms of thought and shows "how this leads on to the elucidation and comparison of different forms of life" (Ibid.).

To summarize Winch's idea, the normativity of meaning lies internally within the practice of its own participants. When it comes to the problem of how social scientists study social phenomena that they themselves are part of, Winch answers that it is conceptually impossible to study social phenomena without being participants in such phenomena. An implication of Winch's idea is not only that criticism of one's community practice is not possible from outside its own conceptual scheme, but also that even within one own community practice, it is not clear how internal norms can be the only source of correctness. The latter brings in the problem quite like what Wittgenstein says, "As if someone were to buy several copies of the morning paper to assure himself that what it said was true." (PI 2009: §265). This is Scylla's situation: the community's norms are so overloaded that it becomes difficult to distinguish between what is right and what appears right. An example might be this. Suppose one lives in a witchcraft culture. It is possible to find certain aspects of life there unsatisfactory, such as the suppression of women, children, animals, the elderly, and disabled people. It seems, according to Winch's analysis, employing the term "suppression" is meaningless because the rule governing the practice of witchcraft culture does not register the signification of the term. People in the witchcraft culture might merely acknowledge terms like "sacrifice" or "immolate". But "suppress" belongs to an outside conceptual scheme and practice. As Winch said,

(...) the relation between idea and context is an internal one. The idea gets its sense from the role it plays in the system. It is nonsensical to take several systems of ideas, find an element in each which can be expressed in the same verbal form, and then claim to have discovered an idea which is common to all the systems. This would be like observing that both the

Aristotelian and Galilean systems of mechanics use a notion force, and concluding that they therefore make use of the same notion. (Ibid.:107)

Even though my example does not concern the same verbal form in different systems of ideas, it reflects that Winch's analysis traps us in the overloaded normative practice where criticism grounded on the external fact about the lives within turns out to be nonsense. But it is absurd to pay no heed to such a fact. I suggest we need the non-intentionally overloaded thick view as a way out.

3. The Non-Intentionally Overloaded Thick View (NTV)

The NTV partly shares Winch's thick view in that it requires the idea of rules engaged within practices. However, Winch's thick view is insufficient to avoid the dilemma of Scylla. What we need is an engagement with rules, the kind of which is objective and not independently characterised from the engagement. Such rules need to be characterized based on regularities of behaviour, but not reducible. However, as we are a part of the behaviour that is an object of our study, whether we are aware of it or not, the notion of practice cannot be a thin notion.

The idea I employ here is Haslanger's view on social practices, which she defines as "... ways of organizing ourselves either towards some end or in response to a coordination or access problem." (Haslanger 2013:10). In other words, practices are patterns of behaviour and normative; they can be right or wrong depending on the purpose we share in living together. Haslanger's notion of practice is a sort of NTV because it is possible that as we repeatedly follow the rules, we are not always aware of which rule we follow. We do it immediately, but we are also aware of what it is to do things correctly or incorrectly.

What keeps practices the way they are is how we interpret each other and the material world, that is, how we posit social meanings under the 'structure' of practices. For Haslanger, social meanings are embedded in our cultures and constrain both individual and collective actions. They are schemas that we learned in language, especially the language of classification, e.g., 'slut', 'mother', and 'woman' (Ibid.). In the case of conflicting practices, the question

is how we know that the social meanings we live with are problematic, so that we need to replace them with other concepts. How do we know that the social meaning of “sacrifice” in a witchcraft culture actually means “suppress” viewing from different conceptual schemes? Haslanger’s answer is

Suppose that through normative inquiry we determine that a particular social practice is misguided and unjust. One question we should ask is how our discursive practice are implicated: how does our language support the categorization that the social practice relies on? Does it prime us to respond in ways that are problematic, e.g., to stigmatize or idealize? (Ibid.: 17)

The answer here might sound circular, for it seems that we need to be in a position outside the practice we are in first in order to recognize the ‘unjust’. However, as Haslanger suggests, social practices are purposeful; for example, in easing our life to live together, one might say that the obstruction of such purpose is an indicator of the recognition of the ‘unjust’. Such an obstruction comes from the way we recognize that social meanings contain an element of stereotypical meanings, i.e. fixed meaning, so that it affects our attitudes and behaviour.

However, according to Haslanger, besides the stereotypical meanings, social meanings also contain their extensional meaning, which tracks truth (Ibid.: 19). For the ameliorativist view, in order to improve the unfairness, social changes require the change of social meanings in both directions – the stereotypical meaning and the extensional meaning. As Haslanger said, “... philosophical inquiry is an inquiry into the concepts we (collectively) ought to use.” (Ibid.: 22). “So we should be asking not simply what concepts track truth, even fundamental truth, but rather: What distinctions and classifications should we use to organize ourselves collectively? What social meanings should we endorse?” (Ibid.: 23)

The last question appears to address the issue of incommensurability that we faced in the earlier discussion on Diamond and Winch. If there are conflicting social meanings or language games, we should choose which one. Which game should we play? What criteria should we adopt for choosing? An answer is already determined by the ameliorative aim. That is to better our collective

life. We need different rules that create different regularities. However, it might be questioned whether this is another Scylla situation where rules are mere interpretations. The question is how the rule that is both objective and not independently characterized from practice can be realized.

The answer might be found in the disjunctive account of direct realism. The element of truth-tracking in social meanings simply grows from the way one 'perceives' how things are disjunctively. Our perceptual experiences are trained by the practice we are familiar with. It is the habit, the regularities of responding to the world in the way we are accustomed to. However, what constrains the experiences is the causal properties of things in the world. It can be taken as the 'phenomenal regularities' or "a regularity of behavior that emerges from the real causal properties of a thing, but that does not itself give rise to or constrain the things' behavior" (Little 1993:187). An example is "glass flows slowly" (ibid.). It is the sort of regularity resulting from the causal properties of glass. It can be lawlike in the counterfactual sense (e.g., if a material is not glass, its physical properties do not flow slowly.); but is not the result of deterministic natural laws. The notion of phenomenal regularities resembles Kripke's notion of a posteriori necessity, which depends on our epistemic abilities. For example, it is metaphysically necessary that in all possible worlds, Aristotle is identical to Aristotle. But epistemically, Aristotle might not be the teacher of Alexander the Great in a possible world. For a natural kind term like 'glass', it is a posteriori necessity that 'glass' is identical to objects that expand slowly because the causal properties of 'glass' might be different in possible worlds.

My point in bringing in the notion of 'phenomenal regularities' is to show how an account of normative but not overloaded practices is possible. If Haslanger is right, then social practices are the means through which social structure exerts its causal power since it has the capacity to produce phenomenal regularities which affect individual and collective action. It is therefore important that we acknowledge the reality of the patterns of behaviour, such as repeated instances of maltreatment behaviours, that individuals experience as the effect of a social structure. This means that the NTV entails realism about social structure.

It may be argued that the NTV faces the circularity problem, that is, behavior regularities are assumed to be the effect of some causal properties of a social structure before it is known whether such properties exist. There might be some kind of patterns or regularities for a certain period, but it is uncertain whether these regularities are necessarily caused by something. However, as mentioned above, this objection is implausible if the regularities in question are taken as phenomenal regularities that can be stated as counterfactual conditions and do not assume deterministic general laws as the cause of regularities.

4. Conclusion

Based on McDowell's reading of Wittgenstein's rule-following argument, the way out of the paradox of rule-following is in a direct grasp of rules without interpretation, that is, in 'habit', 'custom' or 'the meeting of minds' or social practices. I have argued that the most plausible account of social practices is the non-intentionally overloaded thick view because it does not fall into the dilemmatic situation of Scylla and Charybdis. Other types of reading of the rule-following argument, i.e. Winch and Diamond, fail to offer the way out of the dilemma. The NTV is most apt for the ameliorative aim of doing philosophy, although there might be a circularity problem. However, I have argued that the sort of regularities in need is phenomenal regularities, which can avoid such a problem.

References

- Conant, James (1998) "Wittgenstein on Meaning and Use", *Philosophical Investigations* 21, 222-250.
- Diamond, Cora (1991) *The Realistic Spirit: Wittgenstein, Philosophy and the Mind*. Cambridge, MA, USA: MIT Press
- Haslanger, Sally (2013) *Social Meaning and Philosophy Method*. American Philosophical Association 110th Eastern Division Annual Meeting [December 2013].
- Kripke, Saul (1980) *Naming and Necessity*. Oxford: Blackwell.
- Kripke, Saul (1982) *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Little, Daniel (1993) "On the Scope and Limits of Generalizations in the Social Sciences", *Synthese* 97, 183-207.

McDowell, John (1984) "Wittgenstein on following a rule", *Synthese* 58, 325-364.

Sillari, Giacomo (2013) "Rule-following as coordination: a game-theoretic approach", *Synthese* 190, 871-890.

Williams, Bernard (1985) *Ethics and the Limits of Philosophy*, London: Fontana.

Winch, Peter (1958) *The Idea of a Social Science and Its Relation to Philosophy*, London: Routledge&Kegan Paul.

Winch, Peter (1964) "Understanding a Primitive Society", *American Philosophical Quarterly* 1 (4), 307-324.

Wittgenstein, Ludwig (2009) *Philosophical Investigations/ Philosophische Untersuchungen*, Edited by P.M.S.Hacker and Joachim Schulte, translated by GEM Anscombe, PMS Hacker and Joachim Schulte. New York: Wiley.

Wright, Crispin (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.

Signalling Systems: Logic as a Conventional Facet of Reality

Emelia Stanley (Vienna, Austria)

Abstract

Logical conventionalism has been extensively disputed since Quine's (1936) infamous analysis of the position; in particular, it has been charged that there is no notion of convention appropriate to support a logical conventionalist thesis in a sufficient and non-trivial fashion. Here, I sketch the approach to a coherent logical conventionalist thesis by expanding David Lewis' notion of conventionality as signalling equilibrium, and showing how logical inference arises as a natural process of maximising payoff in cooperative signalling games. Drawing from an example by Skyrms, I aim to show how arbitrary signalling games can encode consequence relations that tend towards classicism when reasonable additional constraints are imposed: I conclude that a Lewisian conventionalism about logic can circumvent the Quinean charges.

1. Introduction

Conventionalism about logic, since Quine's infamous (1936) analysis of the position, has often been rejected by philosophers. Aside from the question of the source of our logical conventions (the paper's infamous Regress Argument claims, broadly, that logic is a prerequisite to its own conventional creation) Quine's problem of *what exactly conventions are* also haunts the theory's legacy. Quine was apt to note that some species of conscious agreement, or "explicit stipulation" was highly insufficient to justify logical truth; yet, understood as broadly as "linguistic regularities" or "implicit agreement," conventionalism becomes true in the most vacuous of senses, as our assertions of false superstitions or empirically-known facts equally become "conventional" in origin. While new conventionalist accounts are now promising a path between this Scylla and Charybdis, for instance by delimiting conventions as wholly syntactic or computable rules, here I offer an approach that has not been much explored in the literature, taking conventions as the game theoretic notion of arbitrary stable equilibria following Lewis' (1967) model.

Lewis developed his theory explicitly to assuage Quine's concerns with the convention concept. While not extending his account to logic, Lewis' analysis provides a powerful starting point, when supplemented with further resources, such as Skyrms' observations into signalling equilibria, to characterise conventionalism. Here, I present the skeleton argument for the

conventionality of logic from signalling systems, expanding the coordination equilibrium model of conventions to dissolve the Quinean charges. I show how arbitrary signalling games encode a species of consequence that gives rise to the use of inferential behaviour, and how this consequence can tend towards classicality given certain additional assumptions.

2. An Overview of Signalling Games

A signalling game $\langle S, \Sigma, A, \pi \rangle$ has the following basic setup. One agent, here called the *observer* (or usually the “sender”), detects one of multiple states, $s_1 \dots s_n \in S$. Given s , by some rule they select a signal, $\sigma_1 \dots \sigma_n \in \Sigma$, to send to the *receiver*. Given σ and a rule of their own, the receiver selects an action $a_1 \dots a_n \in A$ to perform. This action has a payoff with respect to the state, $\pi(a|s)$, for sender and receiver, though here we can assume both share a single payoff. Consider a lookout with two bells (σ_1, σ_2) that warns the chef whether an abbot (s_1) or royalty (s_2) is coming, in order to prepare a simple (a_1) or lavish (a_2) meal. We shall represent the observer’s rule or strategy by a function $O : S \rightarrow \Sigma$, and the receiver’s by $R : \Sigma \rightarrow A$. This gives a *strategy* for a signalling game $\langle O, R \rangle_{\langle S, \Sigma, A, \pi \rangle}$.

A signalling game is a cooperative or zero-sum game, in which each agent must select a strategy from a list of options — a function O and a function R — and both benefit from positive payoff where O and R are “calibrated” in a correct way. If neither sender nor receiver can increase their payoff by deviating from their strategy, holding the strategy of the other fixed, the game is in equilibrium and we refer to the strategies as a signalling system. In the example, assuming $\pi(a_n|s_n) = 1$ and 0 otherwise, clearly there are two equilibria: $O(s_n) = \sigma_n, R(\sigma_n) = a_n$ and $O(s_1) = \sigma_2, O(s_2) = \sigma_1, R(\sigma_1) = a_2, R(\sigma_2) = a_1$.

In principle, we could allow for a much richer model by including probabilistic observer and receiver strategies, e.g. where the observer sends signal σ_1 given s_1 only $x\%$ of the time, or receiver performs action a_2 $y\%$ of the time, and so on. Especially where we do not assume a uniform probability distribution across states in S , there can be successful signalling strategies with pooling or partial-pooling equilibria (i.e. where the same signal is sent in response to

multiple states). Such systems are also inevitable where the number of available signals or actions are limited. Here, we will idealise-out such cases, assuming agents have a replete set of symbols and (sometimes) actions, and always develop deterministic strategies for signalling. Later, we will justify these assumptions and make them more precise.

3. The Vervet Monkey Game

In *Signals* (2010), Skyrms provides some simple examples of logical representation that seem to arise in signalling games. His analysis being brief, but along the right lines, in this section we introduce his key example and informal reasoning, before generalising and developing the argument using the notion of an *epistemic scenario* for a signalling game.

Skyrms presents an idealisation of actual game played in nature, where Vervet Monkeys make alarm calls for different approaching predators. We can mirror this example by slightly amending our original dinner party game.

Suppose an observer witnesses one of three guests arriving (s_1, s_2, s_3), sends one of 6 signals, and a receiver prepares one of 6 different meals, $a_1 \dots a_6$. These meals are associated with the guests according to the following payoff table:

	a_1	a_2	a_3	a_4	a_5	a_6
s_1	1	0	0	0.6	0	0.8
s_2	0	1	0	0.6	0.8	0
s_3	0	0	1	0	0.8	0.6

It is clear that, for any optimal signalling system, $R(O(sn)) = a_n$, as the payoffs highlighted bold are the maximum payoff values for that state. This system would disregard the actions a_4, a_5, a_6 .

However, the situation changes if we additionally assume that, occasionally, the observer is unable distinguish between two of the arriving guests. In such a case, the observer is at most aware that one of two different guests is arriving, for instance either *guest 1*, or *guest 2*. We might imagine this new

game to therefore involve 6 states; either the observation of a singleton state such as $\{s1\}$, or of an epistemic scenario, or collection of states in which it equiprobable that one of two states occurs, such as $\{s1, s2\}$. (A singleton is still an epistemic scenario, merely one that corresponds to certainty.) We will assume accordingly that the payoff for an action relative to an epistemic scenario is its average payoff across the states of the scenario. From this, we derive the following payoff table for the (epistemically expanded) dinner party game:

	$a1$	$a2$	$a3$	$a4$	$a5$	$a6$
$s1$	1	0	0	0.6	0	0.8
$s2$	0	1	0	0.6	0.8	0
$s3$	0	0	1	0	0.8	0.6
$\{s1, s2\}$	0.5	0.5	0	0.6	0.4	0.4
$\{s1, s3\}$	0.5	0	0.5	0.3	0.4	0.7
$\{s2, s3\}$	0	0.5	0.5	0.3	0.8	0.3

$a4$, $a5$ and $a6$ are then optimal actions to the three non-singleton epistemic scenarios. Intuitively, if one is unsure whether *guest 1* or *guest 2* is arriving (but is certain *guest 3* has abstained) the best meal to prepare is $a4$, to hedge one's bets for higher payoff. In this sense, Skyrms conjectures the sign $R-1(a4)$ in any signalling equilibrium can be associated with the exclusive disjunction of $s1$ and $s2$ (and, he misses, the negation of $s3$), as we shall make more precise by expanding our framework.

4. Representation in Signalling

We could associate each state s in a state-space S for a game with a truth-assignment $T(s)$ over some variables V . Each epistemic scenario, $\mathfrak{s} \subseteq S$, is then associated with some set of truth-assignments over V , $T(\mathfrak{s})$. In this way, each epistemic scenario is associated with an equivalence class of propositional formulae, $M(\mathfrak{s})$; those formulae composed atomically of variables of V that are

satisfied in only in the truth-assignments $T(s)$. Specifically, for some set of truth-assignments $T(s)$ over variables V , $M(s)$ is $\bigvee_{t \in T(s)} (\oplus tv_1 \wedge \dots \wedge \oplus tv_n)$, where $\oplus tv$ is “ v ” if $t(v) = 1$ and “ $\neg v$ ” if $t(v) = 0$.

By extending payoff from states to epistemic scenarios of S , as before, we can pair each action $a \in A$ with each of those epistemic scenarios for which it is an optimal response. If a is a uniquely optimal response for s , we will say a represents $M(s)$ uniquely.

It is in this way that Skyrms associates a_4 , a_5 and a_6 (or the signals for them) with “exclusive disjunction.” We can easily associate the three possibilities for dinner guests arriving (s_1, s_2, s_3) with three truth-assignments (t_1, t_2, t_3) over three variables (v_1, v_2, v_3). Then the action a_4 is then optimal, uniquely, for the set $\{t_1, t_2\}$, where $t_1(v_1) = 1, t_1(v_2) = 0, t_1(v_3) = 0$ and $t_2(v_1) = 0, t_2(v_2) = 1, t_2(v_3) = 0$. The equivalence class of $((v_1 \vee v_2) \wedge \neg(v_1 \wedge v_2)) \wedge \neg v_3$ (the equivalence class uniquely true for these two truth-assignments over these variables) can therefore be uniquely associated with a_4 , and in this sense a signal for this action can “stand for” the (exclusive) disjunction of two variables (and the negation of a third). Skyrms does not consider the epistemic scenario of total ignorance, $\{s_1, s_2, s_3\}$, but if we add this set, note that this association would “reveal” a_5 to be the action optimal for $\{t_1, t_2, t_3\}$, and therefore to also represent the equivalence class “ $(v_1 \vee v_2 \vee v_3) \wedge \neg(v_1 \wedge v_2) \wedge \neg(v_1 \wedge v_3) \wedge \neg(v_2 \wedge v_3)$.”

However, this imposes the logical structure on the game only insofar as we initially characterise S in a model-theoretic way. Note that we could equally have associated S with three truth-assignments over two variables, and associated a_4 with a completely different equivalence class of formulae. For example, if we associate $\{s_1\}$ with an assignment $t_1(v_1) = 1, t_1(v_2) = 1$ and $\{s_2\}$ with $t_1(v_1) = 1, t_1(v_2) = 0$, then a_4 comes to represent the equivalence class “ v_1 ”, and a_1 the equivalence class “ $v_1 \wedge v_2$ ”.

What is unusual, or even intuitively inappropriate about this mapping of truth-assignments to states is that a_1 and a_4 have no relationship to one another in the game, despite the logical relationship between the sets of truth-assignments these actions stand-for in the mapping (the former entailing the latter). The optimal action for the “ $v_1 \wedge v_2$ ” state is not optimal for the “ v_1 ”

state, though notionally the latter situations should be a subset of the former. This problem did not burden the previous mapping, where the equivalence classes “ $v1$ ”, “ $v2$ ”, “ $v3$ ”, “ $(v1 \vee v2) \wedge \neg(v1 \wedge v2)$ ”, “ $(v2 \vee v3) \wedge \neg(v2 \wedge v3)$ ”, “ $(v1 \vee v3) \wedge \neg(v1 \wedge v3)$ ” were all logically independent of each other.

This suggests a means of introducing further criteria for our mapping between states and truth-assignments: a coherence between the optimality of actions, and the sets of truth-assignments mapped to the states the actions are optimal for, as we shall now develop.

5. Consequence in Signalling

For a signalling game, there is a natural concept for a consequence relation between actions, which we will denote “ $a1 \dots an \models_R a$ ”. Informally, we will say that one action is a consequence of another (or others) iff the consequent action is an optimal choice for the receiver whenever all the antecedent actions are optimal. This mirrors, somewhat, the notion of truth-preservation in logical consequence: one might imagine stating a true proposition as an “optimal response” to some given observed state. We define consequence formally as follows:

$$\Gamma \models_R a \text{ iff, for all } s \in S, \text{ if } \pi(\gamma | s) \geq \pi(a' | s) \text{ for all } a' \in A, \text{ for all } \gamma \in \Gamma, \text{ then } \pi(a | s) \geq \pi(a' | s) \text{ for all } a' \in A.$$

For convenience, we can omit the set notation. It is easy to see that \models_R is a reflexive ($a \models_R a$), transitive ($a \models_R b, b \models_R c \Rightarrow a \models_R c$) and monotonic ($a \models_R b \Rightarrow a, c \models_R b$) relation: broadly, if S_a are those states for which a is an optimal action, then $a \dots an \models_R a'$ iff $S_a \cap \dots \cap S_{an} \subseteq S_{a'}$.

We could also develop a notion of state consequence for the observer, $s1 \dots sn \models_O s'$. Broadly, a state is a consequence of others if all actions that are optimal for the antecedent states are optimal for the consequent states:

$$\Delta \models_S s \text{ iff, for all } a \in A \text{ if } \pi(a | \gamma) \geq \pi(a | s') \text{ for all } s' \in S, \text{ for all } \gamma \in \Gamma, \text{ then } \pi(a | s) \geq \pi(a | s') \text{ for all } s' \in S.$$

We might then conjecture the following rule for adequate mappings of truth-assignments over variables V to a set of states S for a game:

For all $s \subseteq S$, if $M(s_0) \dots M(s_{n-1}) \models_C M(s_n)$, then $a_0 \dots a_{n-1} \models_R a_n$, for each a where a_i is optimal for s_i .

In other words, if $\varphi \models \psi$, then if a is an optimal action for the state associated with φ , and a' is optimal for the state associated with ψ , then we should have $a \models_R a'$. This criterion serves to eliminate the previous pathological mapping for the dinner-party game, as clearly $v_1 \wedge v_2 \models v_1$, but a_4 is not a sender-consequence of a_1 .

6. The Purpose of Inferring

We can see that the consequence of some actions for others represents a kind of domination of one strategy over another: if $a_1 \dots a_n \models_R a'$, then whenever the receiver could use any of the actions $a_1 \dots a_n$ optimally, they could use a' with no sacrifice to payoff. For observer consequence, if $s \models_O s'$, then whenever one could signal for s , no payoff is sacrificed by signalling for s' instead.

This suggests the question of *why* one would move from one strategy to another, as we gain no additional payoff, and could do instead with one fewer signal and action by simply dropping a from the game if $a \models_R a'$. At best, knowing the observer and receiver consequences for a game would seem to show only which signals and actions are eliminable. But by adding two small assumptions to the game, the purpose of inferring to maximising payoffs becomes much clearer.

There are two cases of consequence. For the first, information about the epistemic scenario is effectively decreased, as the antecedent action is more seldomly optimal than the consequent action. This corresponds to moving from an action optimal for an equivalence class such as $p \wedge q$ to the class p , or from p to $p \vee q$.

A decreasing-information inference may be advantageous given the additional assumption that there is some likelihood of an inaccurate signal for the

epistemic state, or signalling noise. There would then be a certain advantage to moving to a more generally optimal strategy, as payoff cannot decrease if the signal is accurate, but in case of inaccuracy the expected payoff is higher.

The second case is an “increasing-information inference.” Suppose we have a_1 , $a_2 \models_R a_3$, where a_3 is an optimal action in fewer epistemic scenarios than both a_1 and a_2 (e.g. this could correspond to an inference from the actions corresponding to “ p ” and “ q ” to action “ $p \wedge q$ ”). This then means changing to an action that is still optimal for the scenario, but is more seldom an optimal strategy throughout the game. We can see how this will be advantageous if we introduce a distinct notion of payoff for the game, called the costed payoff :

The costed payoff of a for scenario s , $\pi_c(a|s) := \pi(a|s) - \pi(a|s')$, where $s' := \{s \in S : s \notin s\}$.

Flat-footedly, the costed payoff adds an additional price to each action, namely, a “opportunity cost” for playing an action that would have had payoff in other states and epistemic scenarios. Costed payoffs would arise, for instance, where we play the game multiple times, but actions have a *limited* number of uses. It is then advantageous to use an action that has the same payoff for a given situation, but more seldom has an optimal payoff for situations in the game. An increasing-information inference necessarily increases the costed payoff of an action, without sacrificing any payoff.

If two agents repeatedly play a noisy signalling game where the receiver has some opportunity cost for actions, then there is an incentive to internalising both types of receiver-consequences for the actions. We may also more simply imagine that, sometimes, certain actions are simply not available to the receiver, so they must substitute for what they have available using the consequences. Ascertaining these consequences requires nothing more than familiarity with the payoffs for each action given each state, or the basic structure of the game; it seems likely that even very simple learning from repeated play would therefore be sufficient for players to determine the consequence relations for their game, and learn to perform a kind of “inference” to maximise the (costed) payoff.

7. Logical Conventionality

So far, we have seen how a notion of consequence between signals and between actions arises for any arbitrary signalling game, and how awareness of these relations can benefit participants. We have begun to introduce, using these consequences, a notion of the representation of equivalence classes of propositional formulae, by epistemic situations of states to sets of truth-assignments over sentential variables. Here, we conclude by sketching a path to a conventionalist account of logical inference, considering how rich logical structure can arise in signalling games and the sense in which this structure can be called conventional.

The games we have considered contain only a minimal amount of logical content, due to a limited number of states and actions that omit and pool equivalence classes. To obtain a full equivalence between logical and receiver consequence over some variables V , we require a game with $2^{|V|}$ states. Moreover, for each $s \subseteq S$, there is an action that is optimal for all $s \in s$. This then ensures that, for each epistemic scenario, there is some $a \in A$ that is the uniquely costedly optimal action for the scenario. Such a game then encodes, we conjecture, a classical propositional consequence relation for $V : a_1 \dots a_{n-1} \models a_n$ iff $M(s_1) \dots M(s_{n-1}) \models M(s_n)$, where a_i is the uniquely costedly optimal action for s_i .

Such a game may arise under the following conditions. Firstly, our states are compositional, each state being the same as the presence or absence of $|V|$ different possible sub-states. For our dinner-party game, we would have to add the states in which none, two guests, or all three guests arrive. Secondly, we can imagine that agents have the ability to develop new actions (and symbols for these actions), perhaps purely stochastically, that are optimal for all states in each epistemic scenario, saturating their options. This then yields a game with a uniquely costedly optimal action for each scenario.

Following Warren (2020: 10), logical conventionalism can be formulated as the position that logical truths are explained wholly conventionally, or employ *only conventions in their explanations*. As presented here, logical consequence is a feature of signalling systems for signalling games, and there are several senses in which signalling systems can be viewed as conventional. Firstly, the signals used in a game are entirely arbitrary, and more importantly, any

uniform permutation of the signals in O and R in a signalling system is also a signalling system, so we always have a plurality of possible conventions. Secondly, signalling systems are able to arise *spontaneously* from simple learning processes, such as Roth–Erev Reinforcement, Bush–Mosteller Reinforcement, exponential response learning, and the “ARP model” (Skyrms 2010: ch. 7). Very succinctly, there is reason to believe that basic learning dynamics, replicable even by unicellular organisms, can lead to stable signalling equilibria in certain conditions, which supports the notion that language and cognition is not a prerequisite to successful signalling, despite anticonventionalist arguments founded on this concern. Developing a replete and classical consequence relation may require some conditions that imply less modest abilities, but we have shown how more restricted forms of nonclassical inferential structure can be embedded in any game given a truth-assignment mapping. This may suggest that weaker forms of inference could arise spontaneously with optimal strategies for signalling games if there is a purpose to inferring. With this, a charge is avoided that the use of reasoning necessitates familiarity with some system of logical reasoning: it might arise as a natural product of an effective signalling strategy.

References

- Lewis, David K. (1969) *Convention: A philosophical study*. Cambridge, MA, USA: Wiley-Blackwell.
- Quine, W. V. O. (1936) "Truth by convention", in: *Philosophical essays for Alfred North Whitehead*, 90–124. London: Longmans, Green & Co.
- Skyrms, Brian. (2010) *Signals: Evolution, learning, and information*, Oxford University Press.
- Warren, Jarred. (2020) *Shadows of syntax: Revitalizing logical and mathematical conventionalism*, New York, USA: Oxford University Press.

Acknowledgement

Thank you to fipWiP seminar attendees for their feedback on the exploratory presentation of this paper.

Betrieb Wittgenstein am Ende doch nur Metaphysik?

Matthias Statzkowski (Berlin, Germany)

Abstract

Bei Philosophen, die über das Philosophieren philosophieren, stellt sich immer die Frage, ob den Worten auch Taten folgten. Wird nicht nur vorgegeben, wie zu philosophieren ist, sondern wird auch genau in diesem Sinne selber reflexiv philosophiert? Wittgenstein war ein sehr differenzierter Denker, aber er wird oft mit einer pluralistischen Position identifiziert, nach der die Realität viele gleichberechtigte Facetten hat. Die Frage ist, inwiefern ein solcher Pluralismus konsistent mit Wittgensteins eigenen Kommentaren zur Philosophie ist. Nach Wittgensteins sonst anti-metaphysischer Haltung sind die Fragen der althergebrachten Metaphysik nämlich problematisch. Für pluralistische Interpretationen droht so eine epistemische Asymmetrie, nach der Wittgenstein härter mit den Metaphysiker*innen ins Gericht geht, als mit sich selbst. Die Behauptung, die Realität hätte unterschiedlich viele Facetten, ist schließlich eine Behauptung, die denen der althergebrachten Metaphysik, was Rechtfertigungsbedarf angeht, in nichts nachsteht.

Auf den Punkt bringt es Ted Sider. Für Sider begeht ein solcher Pluralismus am Ende doch nur Metaphysik: "Many are drawn to metaphysical deflationism because they want the epistemic high ground. They want to rid the world of difficult-to-answer substantive questions. But their very metaphysical deflationism costs them the high ground. For since metametaphysical critiques are just more metaphysics, they raise all the old epistemic questions." (Sider 2011: 83) Ich entwickle in diesem Artikel den Einwand Siders gegen pluralistische Interpretationen von Wittgenstein und stelle schließlich einen konsistenten Wittgenstein vor, der Sider angemessen antworten kann, ohne selber doch nur Metaphysik zu betreiben.

1. Wittgensteins Pluralismus und die Bedrohung einer epistemischen Asymmetrie

Wittgenstein ist bekannt für eine Haltung, die man grob 'anti-metaphysisch' nennen kann; oft wird diese Haltung auch als 'deflationär' oder 'anti-theoretisch' in der Literatur spezifiziert. Welche Bezeichnung die zutreffendste ist, ist zwar ästhetisch interessant, soll aber hier keine große Rolle spielen. Textlich werden im Namen dieser Haltung oft die gleichen Stellen aus Wittgensteins späteren Jahren referiert, wie etwa PI §89-133 oder BT §88. Generell, findet man Evidenz für seine anti-metaphysische Haltung stets, wenn Wittgenstein über das Philosophieren philosophiert. Auch wenn sich die Stärke von Wittgensteins anti-metaphysischer Haltung debattieren lässt, steht es außer Frage, dass Wittgenstein viele Fragen, die althergebracht als metaphysisch zählen, wie über die Natur der Zeit, über Solipsismus, Skeptizismus oder die Existenz Gottes, in irgendeinem Sinne als problematisch sieht. Ob Wittgenstein metaphysische Fragen als problematisch sieht, weil sie offensichtliche Antworten haben, weil sie sinnlos oder, weil sie schlichter

Unsinn sind, mag strittig sein, aber unstrittig ist der problematische Charakter der Fragen für Wittgenstein. Zusätzlich zu dieser anti-metaphysischen Haltung wird Wittgenstein auch oft im Namen eines gewissen Pluralismus herangezogen, in dem Sinne, dass es für ihn nicht so etwas wie 'schlichtweg richtige Begriffe' (PPF 2009: §366) oder den in einem philosophischen Sinne 'objektiv richtigen' (PI 2009: §254, PPF 2009: §340) Blickwinkel gebe. Aus meiner Sicht gibt es eine Spannung zwischen diesen beiden Facetten Wittgensteins, mit denen ich mich im Folgenden befassen möchte.

Unter 'Pluralismus' werde ich in diesem Artikel die Position verstehen, dass unsere Facetten derselben Realität, die Sprachspiele Wittgensteins, in einem objektiven Sinne, gleichberechtigt sind. Pluralismus scheint prima facie eine ansprechende Alternative zu der althergebrachten Metaphysik zu sein, und in der Tat findet man viel textliche Evidenz für eine solche pluralistische Interpretation in Wittgenstein. Anstatt über die subjektunabhängige Welt zu philosophieren, ist das Ziel der Analyse der Mensch selber und seine Praktiken. Grob sieht eine solche pluralistische Interpretation wie folgt aus: Die Bedeutungen von Aussagen ist Sprachspiel abhängig. Es lässt sich nicht sinnvoll, nach der Wahrheit, Bedeutung oder Sinn eines Satzes zu fragen, ohne vorher festzulegen, in welchem Sprachspiel man sich befindet. Hat man aber das Sprachspiel festgelegt, kann man aber determinierte Antworten geben. So lässt sich die Frage nach dem freien Willen eines Menschen einfach beantworten, wenn man sich in einem Gerichtssaal befindet und die Kriterien der Beantwortung einer solchen Frage vorher festgelegt sind. Probleme bekommt man nur, wenn man solche Fragen aus ihrem natürlichen Habitat, aus den Sprachspielen, in denen sie zu Hause sind, entfernt und man trotzdem glaubt, sie haben die gleiche Bedeutung. In einem Gerichtssaal ist es klar, warum nach dem freien Willen und der Autonomie einer Person gefragt wird: Es wird versucht festzustellen, ob die angeklagte Person im Rahmen des Gesetzes schuldig ist. Interessant ist in diesem Sprachspiel beispielsweise, ob die angeklagte Person zu ihrer Handlung gezwungen wurde oder nicht. Nach dieser pluralistischen Interpretation von Wittgenstein wird Metaphysik genau dann betrieben, wenn man eine Reihe von Sprachspielen, die wir mit Wittgenstein die Facetten unserer Realität nennen können, mit Regeln hat und von all diesen Sprachspielen ein Satz aus dem Kontext gerissen und abstrahiert wird. Geht es um die Sprachspiele selbst, haben diese zwar kausale

Ursachen, aber keine tiefere Begründung. (Interpretationen, nach denen unsere Sprachspiele tiefere Gründe haben, gibt es auch. Vgl. dazu Stroud 1965. Diese halte ich aber für unhaltbar im Lichte Wittgensteins anti-metaphysischer Haltung.) Es lässt sich vielleicht wissenschaftlich ermitteln, warum dieses und jenes Sprachspiel sich etabliert hat, aber diese kausale Ermittlung ist keine Rechtfertigung oder Begründung des Sprachspiels (PPF 2009: §365, LFM 1976: 341, OC 1974: §474). Jenes oder solches Sprachspiel mag besser zum Bauen einer Brücke sein, aber kein Sprachspiel ist in einem absoluten Sinne besser als ein anderes. Deswegen sind unsere Sprachspiele auch objektiv gleichberechtigt, obwohl man das eine über das andere aus praktischen Gründen bevorzugen kann. Nach dieser pluralistischen Interpretation verweigert man eine Antwort über 'die Realität' in einem kontextunabhängigen Sinne, und man kann auf paradigmatische metaphysische Fragen mit einem niveauvollen „es kommt darauf an“ antworten.

Ich denke, (Moyal-Sharrock 2013), (Schroeder 2018) oder (Dummet 1959) lassen sich beispielsweise, trotz interessanter Nuancen, so lesen, als befürworten sie eine pluralistische Interpretation Wittgensteins. Ich möchte diese Denker*innen nicht über einen Kamm scheren, und ich bin mir sicher, es gibt relevante Unterschiede zwischen ihnen. Für diesen Artikel ist es dennoch sinnvoll, die Ähnlichkeiten dieser Denker*innen hervorzuheben, da sie mindestens bezüglich dessen, was ein bestimmtes Problem der pluralistischen Interpretation angeht, analog sind.

Wenn man sich über die Facetten der Realität Gedanken macht, kommen für Personen mit pluralistischen Intuitionen sehr schnell schwierige Fragen auf. Wie viele Facetten der Realität gibt es und wie werden sie identifiziert? Sind die Facetten der Realität untereinander strukturiert? Wo hört die eine Facette auf und wo fängt die nächste an? Was macht es den Fall, dass die Realität unterschiedliche Facetten hat? Das Interessante bei all diesen Fragen ist, unabhängig von ihrem Gehalt, dass sie von ihrer Form her den Fragen der Metaphysik sehr ähneln. Was ist der Unterschied zwischen dem Philosophieren über die Grenzen von weltlichen Dingen und den Grenzen von Sprachspielen? Was ist es, dass Fragen über die Natur der Facetten unserer Realität gegenüber Fragen über die Natur der Realität simpliciter unterscheiden? Hat man eine ausgeprägte anti-metaphysische Haltung, sollte

man ein gewisses Unwohlsein über diese Ähnlichkeiten haben.

Es droht eine Asymmetrie epistemischer Standards für pluralistische Interpretationen. Darf man hier mit kritischem Blick Metaphysiker*innen beäugen, wenn man selber solch ähnliche Fragen und Probleme hat? Auf den Punkt bringt es Sider, ein Autor, der sonst nicht weiter von Wittgenstein entfernt sein könnte, mit seinem sogenannten 'just-more-metaphysics' Bedenken.

Many scattershot critics move from the difficulties of metaphysical epistemology to the conclusion that metaphysics is something like meaningless, without realizing how close this comes to assuming a crude form of verificationism. Often these critics have a blind spot: they are verificationist when thinking about metaphysics, but not when thinking about other matters (especially about their own bit of philosophy!). (...)

Details aside, my crucial claim has been that a sufficient condition for substantivity is being cast in joint-carving terms. An important consequence is that metametaphysical critiques are distinctively metaphysical in nature. Whether they are correct is a function of the facts—a function of what joints reality in fact has. One cannot do metametaphysics simply by examining metaphysical language and reasoning. For given the sufficient condition, in order to claim that a question is nonsubstantive, one must claim that it is not cast purely in joint-carving terms, and such a claim cannot be supported solely by reflecting on language and reasoning. (...)

Many are drawn to metaphysical deflationism because they want the epistemic high ground. They want to rid the world of difficult-to-answer substantive questions. But their very metaphysical deflationism costs them the high ground. For since metametaphysical critiques are just more metaphysics, they raise all the old epistemic questions. (Sider 2011: 82-83)

Die Behauptung, die Realität hätte viele Facetten, scheint selbst eine Behauptung zu sein, die metaphysischer nicht sein kann. Der Pluralismus ist zumindest zu der Behauptung verpflichtet, es gebe keine fundamentale Struktur der Welt, denn genau deswegen und genau in diesem Sinne sollen die

Sprachspiele der pluralistischen Interpretation objektiv gleichberechtigt sein. Selbst wenn man die Bezeichnung 'metaphysisch' ablehnt, kann man das Problem epistemisch identifizieren. Wittgenstein ist hochkritisch gegenüber der althergebrachten Metaphysik, aber nach der vorgestellten pluralistischen Interpretation hat Wittgenstein die gleichen epistemischen Probleme, die zu seiner anti-metaphysischen Haltung beitragen, und er benötigt ebensolche Erklärungen, die er kritisch sieht. Warum nämlich Sprachspiele, die Facetten unserer Welt, auf jene und nicht auf solche Weise zu individuieren sind, ist kontrovers, genauso wie die Behauptung, die Welt habe keine fundamentale Struktur. So wie Sider es anmerkt: Die Wahrheit des Pluralismus ist schlicht eine Funktion der Fakten darüber, wie die Welt ist. Aber Aussagen darüber, 'wie die Welt ist' machen Metaphysiker*innen schließlich genauso. Selbstverständlich ist die Freundschaft zwischen Pluralismus und der althergebrachten Metaphysik nichts Verwerfliches. Es gilt nur abzugleichen, ob eine pluralistische Interpretation dann konsistent mit Wittgensteins anti-metaphysischer Haltung ist.

2. Wittgensteins anti-metaphysische Haltung

Zu Wittgensteins anti-metaphysischer Haltung gibt es bereits viel Literatur, und in den Debatten ist zwischen stark deflationären Interpretationen und Interpretationen, die Wittgenstein praktisch zu einem Metaphysiker machen, alles verbreitet. Ich möchte hier meine Interpretation zu BT §88 durch die Linse kontemporärer Metametaphysik vorstellen und anschließend Sider aus der Perspektive Wittgensteins antworten.

In BT §88 schreibt Wittgenstein:

(Das Wort "fundamental" kann auch nichts metalogisches, oder philosophisches

bedeuten, wo es überhaupt eine Bedeutung hat.) (...)

(Alles, was die Philosophie tun kann ist, Götzen zerstören. Und das heißt, keinen neuen

– etwa in der "Abwesenheit eines Götzen" – zu schaffen.) (BT 2005: §88)

Wittgenstein scheint sich an dieser Stelle bereits gegen den Vorwurf eines performativen Widerspruchs zu positionieren. Die 'Abwesenheit eines Götzen'

darf nicht ein neuer Götzen werden. 'Götzen' ist bildlich zu verstehen, aber man kann das Bild durch den vorher genannten Begriff der Fundamentalität ausbuchstabieren. Fundamentalität erfreut sich in den letzten 20 Jahren großer Beliebtheit als Werkzeug in der Metaphysik, aber das Bild der Metaphysik ist immer noch dasselbe, auf welches Wittgenstein reagierte (Ts-309,29): Es gibt ein kleines Inventar von Bauteilen, die in strukturierten Beziehungen zueinander stehen, aus denen die Welt entsteht. Es wird von Metaphysiker*innen behauptet, es gebe metaontologisch relevante Unterschiede, wie sich das Bild ausbuchstabiert. Das metaontologische Schisma besteht hier zwischen Quine und Neo-Aristoteliker*innen. Für Quine (vgl. Quine 1953) existiert lediglich das minimal notwendige Inventar der Welt und nur dieses ist in diesem Sinne auch fundamental. Alle Sätze, die wir trotz der Existenz von so wenig Baustoff für wahr halten wollen, sollen aber dennoch nur durch das Kleinstinventar paraphrasierbar sein, sodass sie auch wahr bleiben. Auf der anderen Seite ist für Neo-Aristoteliker*innen (vgl. Schaffer 2009) Existenz ein 'ontological free lunch'. Wichtig ist für Neo-Aristoteliker*innen, dass alle fundamentalen, das ist nicht-begründeten, Objekte, alles andere in der Welt begründen. Für Neo-Aristoteliker*innen existiert alles, aber nur das Unbegründete ist fundamental. Für Quine existiert nur wenig, aber das, was existiert, ist fundamental. Auch wenn Neo-Aristoteliker*innen und Quine sicher bekräftigen würden, dass diese Unterscheidung substantiell, über die Natur der Existenz, sei, so ist für Personen mit einer ausgeprägten anti-metaphysischen Haltung das zugrunde liegende Bild bei beiden sicher das gleiche. Es gibt bei beiden Strömungen einen kategorischen Unterschied zwischen dem fundamentalen Inventar und dem Rest der Welt. Für beide Strömungen ist nur das fundamentale Inventar der Welt relevant für parsimonische Überlegungen zum Abwiegen unterschiedlicher Theorien (Schaffer nennt das 'Occams Laser'. Vgl. Schaffer 2015: 647). Diese Symmetrie rechtfertigt auch die Zuschreibung von Fundamentalität zu Quine, obwohl er selber den Begriff nicht verwendet hat: Wir können 'fundamental' stipulieren, sodass diese Objekte fundamental sind, dessen Existenz für parsimonische Überlegungen bei intratheoretischem Abwägen relevant sind. Der 'Götzen' in BT §88 sollte in diesem Sinne, als Antwort auf die Frage, was fundamental ist, verstanden werden.

Deswegen kann Wittgenstein, aus meiner Sicht, auch kein Pluralist sein, denn das würde das generelle Bild der althergebrachten Metaphysik akzeptieren, aber schlicht eine andere Antwort liefern. Hier würde aus der 'Abwesenheit eines Götzen' tatsächlich ein neuer Götze erschaffen werden, nämlich jener, dass es kein fundamentales Inventar der Welt und lediglich kausal entstandene Sprachspiele gibt. Das deutet Wittgenstein bereits vorher an, wenn er schreibt, dass wenn 'fundamental' eine Bedeutung hat, es nichts metalogisches oder philosophisches bedeuten kann. Die philosophische Bedeutung von 'fundamental' ist genau die eines ontologischen Bodens, auf dem alles andere steht. Ein wichtiger Unterschied zwischen Wittgenstein selbst und pluralistischen Interpretationen ist hier, dass Wittgenstein Kritik an der Bedeutsamkeit des Wortes 'fundamental', im philosophischen Sinne, übt. Anders als die pluralistische Interpretation es behauptet, hat Wittgenstein nicht nur eine negative Antwort auf die Frage, was nun fundamental ist, sondern er problematisiert radikaler den bestehenden Diskurs. Es lässt sich in diesem Atemzug auch Sider antworten: Wenn Wittgenstein den Diskurs selber und die in dem Diskurs verwendeten Wörter als problematisch sieht, so verneint er nicht nur die Behauptung, es gebe ein fundamentales Inventar der Welt.

'Problematisch' ist aber in der Tat vage. In dem zitierten Paragraph scheint es primär um die Bedeutung des Wortes 'fundamental' zu gehen, aber ich denke, man muss das Problematisieren Wittgensteins holistischer betrachten. Die hier bestehende Gefahr ist es, dass sich metaphysische Debatten schlicht erneut reproduzieren und man beginnt, über die Natur der Bedeutung zu philosophieren. Hat Wittgenstein die richtige Ansicht über die Bedeutung von Wörtern oder nicht? Können Metaphysiker*innen nicht genauso gut Englisch und Deutsch sprechen wie Wittgenstein? Ein Insistieren, es gehe schlicht um die Bedeutung von Wörtern, lenkt einen schnell auf eine falsche Fährte, da Wittgenstein auf die Bedeutung von Wörtern in einem deflationären Kontext eingeht. Wittgensteins Ziel ist es, bestimmte Bilder zu problematisieren und nicht neue Theorien über die Welt aufzustellen. Seine Herangehensweise dafür ist es, die althergebrachte Metaphysik intern zu hinterfragen. Wir kommen, so Wittgenstein, in der Metaphysik schlicht nicht zu Ergebnissen, die uns zufriedenstellen. Das lässt sich anhand des bestehenden Diskurses zwischen Neo-Aristoteliker*innen und Quine über Existenz exemplarisch

darstellen. Warum ist die Debatte über die Natur der Existenz wichtig, wenn doch eine starke Symmetrie beider Positionen zu intratheoretischem Abwägen besteht? Das, worauf es bei sowohl Neo-Aristoteliker*innen als auch Quine ankommt, ist, wie man entscheidet, welche Theorie der Welt von zweien die bessere ist, und beide Richtungen haben doch eine erstaunlich ähnliche Antwort darauf. Für Quine und Neo-Aristoteliker*innen kommt es darauf an, welche Theorie von zweien zu einer kleineren Anzahl von minimal notwendigen Objekten verpflichtet ist. Wie sich Fundamentalität und Existenz bei beiden Strömungen ausbuchstabiert, unterscheidet sich, aber weil sich die beiden Begriffe symmetrisch unterscheiden, bleibt das Endergebnis so unglaublich ähnlich. Aus meiner Perspektive ist die bestehende Symmetrie bei Quine und Neo-Aristoteliker*innen Grund genug, daran zu zweifeln, dass in der Debatte um Existenz mehr als eine Ausdrucksform auf dem Spiel steht. Wittgenstein selber elaboriert diese Methode, wenn er die Idee problematisiert, man bräuchte mentale Objekte als Bedeutung von mentalem Vokabular. Es ist nicht schlichtweg problematisch, dass die Existenz von mentalen Objekten angenommen wird. Sondern, das tatsächliche Problem ist, dass mentale Objekte für Wittgenstein eine 'grammatische Fiktion' sind und trotzdem gedacht wird, es wird eine genuine Erklärung geliefert. Mentale Objekte sind nur eine Redensweise (vgl. MS 124, 6). Ein solches Problematisieren ist möglich, ohne sich selbst der einen oder der anderen Bedeutungstheorie zu verschreiben oder den vollen Sinn der infrage stehenden Aussagen zu akzeptieren. Wittgenstein selber macht diesen Punkt elegant in PI §334 (ein ähnlicher Punkt wird auch in PI §513 gemacht). Dort stellt er sich eine Person vor, die versucht, einen beliebigen Winkel mit Zirkel und Lineal zu dreiteilen, obwohl das unmöglich ist. Wir können in einem gewissen Sinne schon nachvollziehen, was die Person tut, die das Unmögliche versucht. Sonst wäre es Wantzel schwer gefallen, zu beweisen, dass man einen beliebigen Winkel eben nicht mit Zirkel und Lineal dreiteilen kann. Wittgenstein behauptet hier jedoch nicht, die Person, die das Unmögliche versucht, spreche kein Deutsch. Es ist keine semantische Angelegenheit das Vorgehen zu problematisieren. So wie Wittgenstein über die Person reden kann, die versucht einen beliebigen Winkel dreizuteilen, so können wir über Metaphysiker*innen sprechen, auch wenn wir den gesamten Diskurs problematisch finden.

Was ist nun wirklich real für Wittgenstein? Gibt es die Facetten unserer Welt oder gibt es sie nicht? Ich denke, für Wittgenstein hat die Ausdrucksform, über unsere Welt im Sinne von Facetten und Sprachspielen zu reden, Vorteile. Beispielsweise verleitet eine pluralistische Redensart nicht so schnell zu denken, es könne nicht anders sein, aber am Ende ist es schlicht eine Ausdrucksform. Um es mit Wittgenstein zu sagen: „Sage, was du willst, solange dich das nicht verhindert, zu sehen, wie es sich verhält. (Und wenn du das siehst, wirst du Manches nicht sagen.)“ (PI 2009: §79) Nach dieser Interpretation Wittgensteins ist seine Philosophie relativ zu den dominanten Bildern der eigenen Zeit zu verstehen. Es gibt heute wie damals, in der akademischen Welt oder im normalen Leben, Bilder, die uns in ihrem Bann gefangen halten und glauben lassen, unsere Begriffe könnten nicht anders sein. Man denke nur an die öffentliche 'Debatte' darüber, wer als Mann oder Frau zählt. In dieser werden oft haltlose, nahezu scholastische, Annahmen getroffen, nur weil man es sich nicht vorstellen kann, dass unsere Genderbegriffe anders sein könnten, als manche Leute glauben. Behauptungen über die Facetten unserer Realität sind so keine quasi-wissenschaftlichen Thesen über die Welt. Vielmehr zeigen sich die Facetten unserer Welt als Ausdruck einer a posteriori Erkenntnis, der Untersuchungen bestimmter Bilder und Begriffe. Die Pointe Wittgensteins Methodologie ist es, die Bilder der eigenen Zeit als Bilder zu offenbaren und nicht einen neuen Götzen zu erschaffen, selbst wenn dieser lediglich die Abwesenheit eines Götzen sei.

Wo Sider recht hat, und wo man als Person mit Wittgensteinschen Tendenzen Obacht hüten muss, ist es Wittgensteins radikalen Charakter nicht zu übersehen, wenn man selber philosophieren möchte, denn genau dann droht die epistemische Asymmetrie, mit anderen strenger umzugehen als mit sich selbst.

References

Dummett, Michael (1959) "Wittgenstein's Philosophy of Mathematics", *Philosophical Review* 58, 324-348.

Moyal-Sharrock, Daniele (2013) "Realism, but not Empiricism: Wittgenstein versus Searle", in: Timothy P. Racine und Kathleen L. Slaney (eds.) *Wittgensteinian Perspective of the Use of Conceptual Analysis in Psychology*, London: Palgrave Macmillan 153-171.

Quine, W. V. (1953) "On what there is" In *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press, 1-19.

Schaffer, Jonathan (2009) "On what grounds what", in: David Manley, David J. Chalmers und Ryan Wasserman (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford University Press 347-383.

Schaffer, Jonathan (2015) "What to Not Multiply Without Necessity", *Australasian Journal of Philosophy* 93, 644–64.

Schroeder, Severin (2018) "On some standard objections to mathematical conventionalism", *Belgrade Philosophical Annual* 30, 83-98.

Sider, Theodore (2011) *Writing the Book of the World*, Oxford, England: Oxford University Press.

Stroud, Barry (1965) "Wittgenstein and Logical Necessity", *Philosophical Review* 74, 504-518.

PPF 2009

Wittgenstein, Ludwig (2009) *Philosophical Investigations / Philosophische Untersuchungen*. Ed. by P. M. S. Hacker and Joachim Schulte, translated by G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. New York: Wiley.

Wittgenstein, Ludwig (1976) *Wittgenstein's Lectures on the Foundations of Mathematics: Cambridge, 1939*. Hg. von Cora Diamond. Ithaca, USA: Cornell University Press.

Wittgenstein, Ludwig (1974) *On Certainty / Über Gewißheit*. Ed. by G. E. M. Anscombe and G. H. von Wright, translated by Denis Paul and G. E. M. Anscombe. Reprinted with corrections and indices. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (2004) *The Big Typescript: TS 213*. Hg. und übersetzt von C. Grant Luckhardt und Maximilian A. E. Aue. Malden (Mass.), Oxford, Carlton (Vic): Blackwell.

Somewhere Between Eden and the Matrix: Variations on Themes from David Chalmers

Twan Stiegel (Berlin, Germany)

Abstract

This paper presents some variations on themes from David Chalmers's philosophy, specifically those of semantic theory and metaontology, the story of Eden, and the possibility of Cartesian scenarios. Building on Chalmers's metametaphysical insight that answers to questions about the nature of reality in general depend on one's semantic theory and on one's metaontology, I propose a minimal metaontology of external objects, on which external objects are characterised as having a spatiotemporal position and as being mind-independent. Believing Chalmers to be right to call the notion of mind-independence obscure, I try to spell out this notion and present a variation on Chalmers version the story of the Garden of Eden in doing so. Lastly, I modify an argument from Chalmers to argue that although radical deception is possible according to the metaontology I propose, this doesn't imply Cartesian scepticism because the possibility of being in a Cartesian scenario doesn't undermine our epistemic practices. I argue that even if we could be living anywhere between Eden and the Matrix, this fact shouldn't worry us very much.

1. Semantic theory and metaontology

In a number of influential papers and a recent book (Chalmers 2005, 2006, 2018, 2022), David Chalmers highlights a central tension in our picture(s) of reality. On the one hand, there is the Edenic picture according to which we are directly acquainted with objects in the world and with their properties. Chalmers calls this picture Edenic because, as he tells the story, in the Garden of Eden, everything was exactly as it appeared; but as we ate from the Tree of Illusion and the Tree of Science, we fell from Eden and this perfectly transparent reality was no more. Only after the fall could we have come up with the opposite picture: that we might be radically deceived about the nature of reality, as different Cartesian scenarios illustrate. The movie *The Matrix*, in which humans are unwittingly emerged in a perfect computer simulation, is discussed by Chalmers as an example of a Cartesian scenario.

In his work on perception and his work on Cartesian scepticism, Chalmers validates both these pictures. On the one hand, Chalmers argues that Eden "acts as a sort of ideal that regulates the content of our perceptual experience" (Chalmers 2005: 50). On the other, Chalmers argues that we might indeed be in a perfect and permanent computer simulation like in *The Matrix*. However, if we were living in a computer simulation, according to Chalmers,

most of our ordinary beliefs about reality would still be true. For his position, which he calls the “no-illusions view of Cartesian scenarios,” Chalmers relies on two things: a semantic theory on which there are some externalist and some internalist words so we can consistently spell out Cartesian scenarios in spite of Putnamian concerns; and a structuralist metaontology on which for something to be an x is for it to have the structure of an x . Chalmers then argues that if something is real in a nonsimulated world, it will also be real in a perfectly simulated world, because it will have the same structure.

I believe there are problems with Chalmers’s semantic theory and metaontology, most notably the following: if referents are fixed by structure and structure is mind-independent, this raises the exact sceptical concerns the “no-illusions view of Cartesian scenarios” sought to avoid because the world might still have a radically different structure than the one it appears to have. Chalmers appears to anticipate this objection when arguing that, on his structuralist view, entities are defined by their mathematical *and observational roles* but I don’t believe this answers the objection because observation is fallible. If unreal things can fulfill the aforementioned observational roles *sometimes*, we’re owed an account as to why they could not do so *always*. As far as I can tell, Chalmers doesn’t provide such an account. His structuralism in itself doesn’t show that radical deception is impossible (nor does it have to, I will argue, as the possibility of radical deception doesn’t necessarily lead to Cartesian scepticism; I will return to this point in section 3).

However, I believe that the most valuable lesson to be learned from Chalmers’s work on Cartesian scepticism isn’t the specifics of his semantic theory or his metaontology, but the metametaphysical insight that answers to questions about the nature of reality in general depend on one’s semantic theory and metaontology. For my own account of reality and Cartesian scenarios, I therefore presuppose a different semantics and metaontology than Chalmers.

With regards to semantics: I only make two minimal assumptions about language. One, I assume that for something to be correctly described as a real x , it must be possible for it to feature in our linguistic practices (so systematically hidden structure can’t fix reference). Two, I presuppose that we

can meaningfully formulate Cartesian scenarios in language (which is intuitive but would arguably be impossible if externalist semantics were true; see Putnam 1981, Button 2013).

With regards metaontology, I defend a minimal notion of external objects on which external objects have two essential characteristics:

- A spatiotemporal position
- Mind-independence

I take this characterisation to be relatively uncontroversial. Spatiotemporal positions can be thought of purely in relational terms: an object may exist at a particular distance from another and at an earlier or later time. Therefore, we don't need to make any assumptions about the nature of time and space. A spatiotemporal position distinguishes from other arguably mind-independent objects, such as abstract objects.

The second characteristic, mind-independence, is one that's widely accepted; however as Chalmers notes, the notion of mind-independence is obscure (2009: 92, footnote 9). In the next section, I will attempt to spell out this notion.

Against this background of this specific semantic theory and metaontology, the following question arises: If we assume that existential judgements are necessarily connected to our linguistic practices, how could we have a concept of reality as being potentially so radically detached from these practices as to disallow epistemic access to the external world, as the possibility of Cartesian scenarios seems to suggest? In the next section, I will attempt to explain this fact by showing that, even if we assume such a necessary connection between existence and our linguistic practices, the world can be very different from how we take it to be.

2. Mind-Independence and the Garden of Eden

Amie Thomasson (2020) makes a distinction between nomological and metaphysical modal claims, where modal claims “licence inferences based on empirical evidence” and metaphysical modal claims don't. As Thomasson makes the point, “the distinctly *metaphysical* modal features at issue in

characteristic metaphysical debates are cases in which we have *the very same empirical information, and same physical laws and properties*, and yet come to different modal conclusions” (155). It’s tempting to think of “There are mind-independent external objects” as a metaphysical modal claim: clearly, to say that something is mind-independent is to make the modal claim that it couldn’t exist without mind; and clearly, realists and idealists are working with the same empirical information. However, I will argue that “There are mind-independent external objects” is a nomological modal claim rather than a metaphysical modal claim. In other words, on my account, whether the external world is mind-independent is an empirical question.

Mind-independence is philosophical notion that isn’t used in ordinary language, but if we are to make sense of it, I believe we must consider what role the notion plays for everyday concerns. Campbell (2014) gives us the following account, arguing that “We can think of the mind-independence of ordinary concrete objects as showing up in three different types of reasoning” (34):

- One and the same object can be encountered at different times
- One and the same object can be encountered in different sensory modalities
- One and the same object can be encountered by different subjects

(Note that real external objects can in principle be encountered by every possible subject. The claim that external objects can be encountered by different subjects should be read in this stronger sense. The ‘in principle’ requires spelling out, but that falls outside the scope of this paper.)

There are least two ways to read the abovementioned claims and I believe that both of them are correct. On the first reading, they are to be interpreted as norms governing our epistemic practices: we labour under the assumption that it’s possible that the same object is encountered at different times, in different sensory modalities, and by different subjects. On the second reading, these are empirical claims: for any observed object, the question of its mind-independence depends on whether these claims are true of it; and whether they are true of it is determined on the basis of observation.

If this is right, we can tell a different version of Chalmer’s story about the garden of Eden. The story goes like this: In Eden, everything was exactly as it appeared. External objects appeared to be mind-independent in that it appeared as if the same object could be encountered over time, perceived in different sensory modalities, and could be perceived by different subjects; and since we lived in Eden, all of this was true. However, as we ate from the Tree of Illusion and the Tree of Science, we learned that appearances could be deceptive. Thus we have fallen from Eden, but Eden still regulates our epistemic practices in two ways. One, we labour under the assumption that things may sometimes be as they appear – that is, mind-independent. Two, mind-independence, according to the criteria laid out above, acts as an ideal to which real external objects must conform. If this story about Eden is true in spirit, as I believe it is, whether the external world is mind-independent is an empirical question.

So why would anybody think that whether there are mind-independent external objects *isn’t* an empirical question? I believe there are two main reasons. First, because our epistemic practices are guided by the Edenic ideal, there’s a distinction between the world as it appears in ordinary perception and scientific objectivity. For example, the third criterion mentioned above effectively states that external objects can in principle be encountered by every possible subject, but most qualities of external objects aren’t universally intersubjective. (I would argue that only *relations* between things can be construed as perfectly intersubjective, making relational structure in some sense ontologically fundamental in our scientific picture of the world, but defending this claim falls outside the scope of the present paper; see Ladyman and Ross (2007) for a proposal along these lines). This implies that the world as it appears in perception is in many respects different from the real external world and creates the illusion of a “veil of perception” behind which the external world lies hidden. However, on my account, the veil is truly illusory as mind-independent reality simply is what’s observed stripped from that which doesn’t pass the intersubjectivity test. This solves the puzzle mentioned at the end of the previous section, as it explains how mind-independent reality can be both inherently tied up with our epistemic practices and at the same time so distinct from subjective perception as to be justifiably called a “view from nowhere.”

The second reason one might think that whether the external world is mind-independent isn't an empirical question builds on the possibility of being in a Cartesian scenario. If I were in such a scenario, observation wouldn't do me any good; thus, if it's possible for me to be in a Cartesian scenario, whether the world exists mind-independently can't be an empirical question. However, this conclusion simply doesn't follow. The possibility of being in a Cartesian scenario merely indicates that observation is fallible, not that observation isn't the right approach when it comes to questions of mind-independence. I will return to this point in the next section, in which I'll discuss the supposed threat of Cartesian scepticism.

3. Even if Cartesian scenarios are possible, Cartesian scepticism is false

I presupposed that we can meaningfully formulate Cartesian scenarios in language. Now I'll go one step further and say, for the sake of argument, that Cartesian scenarios are possible and that Chalmers is right that we could be dreaming a lifelong dream or be living in a perfect, permanent computer simulation. If this were true, would this lead to Cartesian scepticism?

According to Chalmers, it doesn't. He presents the Master Argument of Cartesian scepticism as follows:

1. You can't know you're not in a Cartesian scenario.
2. If you can't know you're not in a Cartesian scenario, you can't know anything about reality.
3. So: You can't know anything about reality.

Chalmers accepts the first but rejects the second premiss on the basis of his metaontology. According to his structuralism, many entities are defined by their mathematical and observational roles:

[P]hotons and quarks are defined in physics by the mathematical roles they play and by their connections to observations. Photons are what play the photon role. Quarks are what play the quark role. It follows that if there's something in the world that plays the mathematical and observational role of a photon, that thing is a photon. If something plays the role of a quark, that thing is a quark.

The second premiss of the Master Argument is therefore false: even if you were in a lifelong dream or in a perfect permanent computer simulation, you could know about photons and quarks. The conclusion to the argument doesn't follow because you could know about reality even if you were in a Cartesian scenario.

On Chalmers's account, if you were to suddenly wake up from a lifelong dream, you should say that most things in your dream were real. On the metaontological account I present, this clearly isn't the case: if you were to wake up from a lifelong dream, you would recognise that the objects you had engaged with up until that point aren't mind-independent and therefore not real. Does this mean that the Cartesian sceptic is right? Is it impossible to know anything about reality and more importantly, does this make our epistemic practices futile?

I don't think so. I think it's a mistake to think that the possibility of radical deception demonstrates the futility of our epistemic practices. Chalmers again makes a good observation about this. He imagines a different kind of Cartesian scenario, one in which an evil demon prevents us from reasoning well:

Perhaps we can't rule out that an evil demon is tampering with our mind, but we can still rule out, by good reasoning, that two plus three is six. Admittedly, someone whose reasoning was twisted by the demon might say there was proof that two plus three is six. They're reasoning badly, so they don't know the truth. But when we reason well, we can know. The same goes for knowledge of the external world. If good reasoning supports the conclusion that there is a chair in front of me, then the mere possibility of this demon doesn't undermine this reasoning. (Chalmers 2022: 452-453)

As Chalmers points out, the conclusion the sceptic wants to draw simply doesn't follow: even if it's possible that we always reason badly, this doesn't mean that we shouldn't try to reason well. Similarly, if the question whether there are mind-independent objects is an empirical question, as I have argued, the possibility that our observations are systematically wrong doesn't imply that we should give up making observations. In this particular case, the possibility of radical deception only indicates nothing more than the fallibility of observation, which is widely understood to be central to our observational

practices. The possibility of radical deception in general, then, doesn't necessarily undermine our epistemic practices or lead to epistemic paralysis as the sceptic would have us believe.

The conclusion we should draw for our our picture of reality is an interesting one, I think. We are probably not in Eden. We might even be in the Matrix. However, in the final analysis, this needn't worry us very much. In determining whether things around us are real, we will inevitably rely on observation. Observation is fallible and could even be structurally get things wrong, but this doesn't mean that should quit. We should just keep calm and carry on.

References

- Button, Tim (2013) *The Limits of Realism*, Oxford: Oxford University Press.
- Campbell, John (2014) "A Straightforward Solution to Berkeley's Puzzle", in: John Campbell & Quassim Cassam, *Berkeley's Puzzle: What Does Experience Teach Us?* Oxford: Oxford University Press.
- Chalmers, David (2005) "The Matrix as Metaphysics", in: C. Grau (ed), *Philosophers Explore the Matrix*. Oxford: Oxford University Press.
- Chalmers, David (2006) "Perception and the Fall from Eden", in: T. Gendler & J. Hawthorne (eds), *Perceptual Experience*, Oxford: Oxford University Press.
- Chalmers, David (2009) "Ontological Anti-Realism", in: D. Chalmers, D. Manley, and R. Wasserman (eds), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press.
- Chalmers, David (2018) "Structuralism as a Response to Skepticism", *Journal of Philosophy* 115 (12), 625-660.
- Chalmers, David (2022) *Reality+: Virtual Worlds and the Problems of Philosophy*, London: Penguin Books.
- Putnam, Hilary (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.
- Ladyman, James; Don Ross (2007) *Every Thing Must Go: Metaphysics Naturalized*, with David Spurrett and John Collier, Oxford: Oxford University Press.
- Thomasson, Amie (2020) *Norms and Necessity*, Oxford: Oxford University Press.

Acknowledgement

This work was supported by a grant from the German Academic Exchange Service (DAAD). Thanks to Niklas Dahl, Alex Murphy, Krzysztof Sękowski, Amie Thomasson and the attendants to the work-in-progress lunch meeting at Dartmouth College for their comments.

The Part of the *Tractatus* That Anscombe Thought Is “Most Obviously Wrong”

Inbal Tamir (Tel Aviv, Israel)

Abstract

In her Introduction to Wittgenstein’s *Tractatus*, Anscombe discusses Wittgenstein’s early remarks on will and action and declares: “It is this part of the *Tractatus* that seems to me most obviously wrong”. The paper offers a more charitable reading of those parts from the TLP, by pointing at relevant discussions of those matters in Wittgenstein’s “Notebooks 1914-1916”.

Anscombe’s concern seems to be that the *Tractatus* fails to give a good account of (a) the relation of the will and our actual actions in the world, and of (b) our actions as related to ethical notions such as good and bad.

The relation of will and action is portrayed in the notebooks such that the Will (*Der Wille*) accompanies the act, compared with mere wishing (*Wünschen*) which precedes it. The paper brings forward the importance of this distinction, and tries to unpack what such “accompanying” might be.

While it might be the case that the *Tractatus* does not let us speak directly of good and bad as attributes of specific actions, actions can still be thought of together with ethical notions such as those. I suggest that Wittgenstein offers to see actions as manifestations of an attitude towards the world, which itself can be considered good or bad. What lies at the heart of early Wittgenstein’s ethical thought is the conception of such an attitude, expressed in the Notebooks entry by which “in order to live happily I must be in agreement with the world”, which I try to clarify.

1. Anscombe’s Critique

Anscombe presents in her introduction to the *Tractatus* an intriguing critique of its parts regarding action and ethics. Anscombe’s critique begins by referring to sections 6.373-6.374 of the TLP: “Wittgenstein insists that ‘the world is independent of my will; there is no logical connection between will and world [...] no logical connection between my will and what actually happens at all. [...] Therefore ‘action’, in the ethical sense, is something independent of what happens; and this is the bearer of good and evil” (Anscombe 1959: 171)

To this, she replies sharply: “It is this part of the *Tractatus* that seems to me most obviously wrong. As Wittgenstein asks in *Philosophical Investigations* (§644): ‘Did not your intention [of which you are ashamed] reside also in what you did?’ ‘What happens’ includes ‘actions’, in the sense of the word in which ‘good’ and ‘bad’ are predicated of actions. But the philosophy of the *Tractatus*

could not allow this to be so; hence the chimerical ‘will’ which effects nothing in the world, but only alters the ‘limits’ of the world.” (Anscombe 1959: 171-72)

Anscombe, as I read her, is concerned with two issues, both express a concern that Wittgenstein’s notion of “the Will” is incompatible with our common notions of action and ethics. The two can be divided as follows:

1. There should be some connection between the notion of “The Will” and specific actions, i.e., actions done by somebody at a certain place and time (which is to say - actions that can be described as part of “what happens”).
2. Such actions are actions “in the sense of the word in which ‘good’ and ‘bad’ are predicated to actions”. Anscombe is therefore bothered by the fact that evaluative (ethical) terms such as those cannot be ascribed to specific actions, given Wittgenstein’s view.

I will approach those two issues in order, but first, a few general words on the notion of “the Will” in early Wittgenstein are in order.

2. Early Wittgenstein on the Will

There are three characteristics to the notion of the Will in the Notebooks which are of extreme importance:

- a. The Will is an attitude of the subject to the world (TB 1961: 87).
- b. The Will is closely connected with the notion of Action, as reflected most clearly by a Notebook entry: “If the will has to have an object in the world, the object can be the intended action itself. And the will does have to have an object” (TB 1961: 87).
- c. The Will is what Wittgenstein calls “the bearer of good and evil” (TB 1961: 76), a notion which then appears in the TLP such that the will is “the bearer of the ethical” (TLP 1922: 6.423).

At this point we should pay attention to a crucial distinction Wittgenstein makes between two notions, a distinction which I think Anscombe fails to give the right emphasis to. This is the difference between “The Will” (*Der Wille*), and “The Wish” (*Der Wunsche*), and the parallel distinction between “Willing” (

Wollen) and “Wishing” (*Wünschen*). This distinction is presented most clearly when the two are brought together in the notebooks, for example: “Wishing is not acting. But willing is acting” (TB 1961: 88), and “The wish precedes the event, the will accompanies it” (TB 1961: 88).

Wishing, as I read it, is the desire that some factual change in the world will happen, that some state of affairs will take place. This of course can involve myself, i.e., a desire for some factual change regarding me, including not only a desire for things happening to me, but also a desire that I will do such-and-such.

This, in turn, cannot be equivalent to willing, given characteristic c of the notion of will mentioned above, according to which the will is “the bearer of good and evil”, or “the bearer of the ethical”. This is so since being the bearer of ethical value, in Early Wittgenstein, is incompatible with the empirical, worldly considerations, reflected in wanting something to be the factual case in the world. Such incompatibility is the result of Wittgenstein’s position according to which everything that is factual is also contingent and accidental, which is reflected most accurately in (TLP 1921: 6.41): “[...] all happenings and being-so is accidental. What makes it non-accidental cannot lie in the world, for otherwise this would again be accidental”.

But of course, we want (and it seems that Wittgenstein wants as well) that value, the ethical, good and bad, will be non-accidental. In fact, their being non-accidental seems to be an ontological characterization of the exact things which they are. At the same time, we want them to be related to action (see characteristic b of the will brought above). Those two demands appear to be contradictory, which is exactly Anscombe’s first problem, to which I now turn.

3. Anscombe’s First Challenge: The Will and Specific Actions

In order to grasp the depth of the problem we are in, it should be noted that this problem is not only that the relation between my actions and their consequences is accidental, and therefore such relation cannot be the basis of ethics. This is not only a critique of what is known as “consequentialism”. The deeper problem is the arbitrariness of any relation between my will and my actions, since the causal relation itself is already too empirical and “worldly”,

and as such cannot be applied to what is the bearer of value, of the ethical, of good and evil.

What maintains a causal relation with actions is the Wish, not the Will. There is, however, a connection between Will and actions. Anscombe’s first problem is the alleged lack of connection between the Will and specific actions, but it should be noted that Wittgenstein specifically points at such connection: “The wish precedes the event, the will accompanies it”. (TB 1961: 88).

The question, however, is what such relation of “accompaniment” amounts to. Given the tension I formulated above, it is now clear that any such relation of the Will and specific actions cannot be described along any causal, empirical, or other worldly lines. Such accompanying cannot just mean “being side by side”. This is so because such relations are part of what happens in the world, and as such, cannot describe the relation between action and what is of value, what is the bearer of ethics. The way I read the relevant parts in the notebooks is that “the Will” should be thought of as an aspect of action, an aspect that constitutes its ethical value. This reading is inspired by an interpretation put forward by Eli Friedlander, according to which the will “permutes action, constitutes so to speak its substance” (Friedlander 2018: 111). Such reading, by all means, calls for further elaboration. Allow me in the current paper to suggest just those preliminary remarks, and move on to Anscombe’s second challenge.

4. Anscombe’s Second Challenge: Actions and Ethical Evaluation

Anscombe’s second challenge concerns the alleged lack of connection between the notion of The Will and ethical evaluation of specific actions. To approach this we should first pay attention to what Wittgenstein writes about the notion of “the happy life”. The pair happy-unhappy, with regards to the notion of “Life”, is closely connected in that period of Wittgenstein with the pair good-bad (“simply the happy life is good, the unhappy bad” (TB 1961: 78)). This of course calls for further research, but for the time being I shall use the two interchangeably.

A happy life, according to Wittgenstein, is life of what he calls agreement with the world: “In order to live happily I must be in agreement with the world. And that is what ‘being happy’ means” (TB 1961: 75).

As I suggest reading the notion of “agreement with the world” in the notebooks, “happy life” is achieved regardless of the factual happenings in the world, regardless of this-or-that being the state of affairs (“The only life that is happy is the life that can renounce the amenities of the world. To it the amenities of the world are so many graces of fate” (TB 1961: 81)). What is crucial here is not to think of that suggestion as a suggestion to be satisfied with “what you have”, since what you have is also a contingent state of affairs, and as such cannot be the source of value or happiness. Rather, the saying that a good life is a life of agreement with the world is pointing to structural characteristics of the notions of value (that is, of “the good”) and of “the world”. Value cannot, by definition, be sought after in the world, in this-or-that being the case, since those are contingent and accidental, and value has to be something non-accidental. Wittgenstein here therefore does not suggest some trivial self-advice, according to which when faced with the miseries of the world you should focus on the glass half full, or that in the case of amenities we should remember that other things are more important than those. Value, good or bad, simply cannot come from the merely factual, the empirical.

But given that value cannot come from within the world, what are we to do with the fact that we, as “willing subjects”, are located in the world? If “good and evil” cannot come from anything in that world, it seems that our life should be described as neutral with regards to ethical value. How are we to talk of “good and bad” at all? (and another way to formulate this thought would be: given that we can, in fact, characterize our life as having to do with value, how is this possible? is this mere illusion?).

The answer, I suggest, comes through reminding ourselves that as human beings we act in the world, that is, through the notion of Action. As Wittgenstein portrays it, the subject is the willing subject. Given his characterization of the notion of the will (see characteristic b of the notion of “the Will” above), that means she is the acting subject.

It is at this point that the distinction between wishing and willing becomes most relevant. As put earlier, wishing is the aspect of acting that is related to the contingencies of the factual. Willing is the aspect of acting that has to do with its having value. The notion of “agreement with the world” as the good life, I suggest, can and should be unpacked along the lines of that distinction.

Martin Stokhof suggests, in his book on early Wittgenstein, that a solution comes through abandoning wishing altogether (Stokhof 2002: 215-17). This, I think, is a problematic suggestion, mainly because what Wittgenstein looks for is agreement with the world, in all its factual contingency. Such a world includes human wishing as part of it, and merely abandoning wishing would render it, simply put, a different world.

But another solution (and here I am again heavily indebted to Eli Friedlander in his reading of early Wittgenstein) is trying to achieve a new understanding of the relation between Willing and Wishing in human action. According to that solution, the good life, as the life of agreement with the world, is life in which willing and wishing merge. More specifically, it is life in which wishing merges into willing: a life in which there is no superfluous expectation to a change in the factual world, which is not already encompassed in one’s action itself; a life in which the gap between what I want to happen, and how I willfully act, disappears. Importantly, this should not be taken to mean that I am happy with everything that happens, but rather that my wishing and willing becomes one, a life in which wishing merges with willing, through action.

How do we get from this to ethical evaluation of specific actions, as opposed to some general attitude towards the world and life (see characteristic a of the Will)? First, It is important not to flatten what Wittgenstein has in mind here and reduce it to other known ethical approaches. The Will is an attitude of the subject towards the world, which means the point of reference is the world in its entirety. However, such an attitude can be, and is, manifested in specific actions. When asked how exactly, we should first remind ourselves that as human beings we act in the world, and therefore, how can such an attitude towards the world come to fruition, can become manifest, if not by our actual action in the world?

Wittgenstein, as I suggest reading him, mentions at least two ways in which this actually happens. First, there is the famous line from the *Tractatus*: “If good or bad willing changes the world, it can only change the limits of the world, not the facts. [...] [the world] must so to speak wax or wane as a whole.” (TLP 1922: 6.43). Now, it is interesting that in the *Notebooks* the sentence ends differently: “The world must, so to speak, wax or wane as a whole. As if by accession or loss of meaning.” (TB 1961: 73). It seems that as long as the facts (“what happens”) includes my actions, it is through good willing, through the attitude of agreement with the world, that those actions can become meaningful or meaningless.

Secondly, and most important, we should take a close look on (TLP 1922: 6.422):

When an ethical law of the form, ‘Thou shalt . . .’, is laid down, one’s first thought is, ‘And what if I do not do it?’

It is clear, however, that ethics has nothing to do with punishment and reward in the usual sense of the terms.

So our question about the consequences of an action must be unimportant. At least those consequences should not be events. For there must be something right about the question we posed. There must indeed be some kind of ethical reward and ethical punishment, but they must reside in the action itself. (And it is also clear that the reward must be something pleasant and the punishment something unpleasant.)”

Let’s see exactly what is going on here: Eventual consequences of an action must be unimportant. That is so given the view that those are contingent and what is truly valuable has to be non-accidental. However, Wittgenstein does admit that “there must be something right about the question”, there must be something right about our searching for a relation between what we actually do and the realms of ethics and happiness.

I suggest reading the saying that “there must indeed be some kind of ethical reward and ethical punishment, but they must reside in the action itself” as saying that there is happiness and unhappiness, good and bad, which are related to actions. Given the relation of willing and happiness, I suggest

Wittgenstein means that there should be some connection between the notions of good and evil and the actual courses of action one takes, which involve specific actions. But this has to be involved already in the action itself, and has nothing to do with its factual consequences, nor with the specific intentions to a factual occurrence which precede our action or lead to it in a causal manner. Therefore, a specific course of action can be thought of as a manifestation of, and a reflection of, good will. That is, of the correct relations of willing subject and world, of “agreement with the world”. In that sense, we could say that some “good” resides in an action itself, as that action is a reflection and a fulfillment of good willing, that is, of life of agreement with the world. The value residing in the action itself resides in it because the action itself is a manifestation of what it is to act in a valuable way.

What about the pleasantness of the reward and the unpleasantness of punishment? We can think of the suggestion here as saying that the person who is acting in good will, that is, is in agreement with the world, is already being rewarded, exactly through his activity in that manner. Conversely, we can think of a person acting out of bad will, who is not in agreement with the world, as being already punished through that precise way of acting. The former’s precise act manifests agreement of subject and world, which is already a pleasant reward, while the latter’s precise act itself manifests disagreement of subject and world, which is itself already an unpleasant punishment. It should not be thought as if the “result” of that action is agreement or disagreement with the world, for the action itself is a manifestation of such an attitude of agreement or disagreement. As such, that precise act is already pleasant or unpleasant.

5. A Concluding Remark

I think that the question whether Wittgenstein’s move outlined here is appealing or not, is another question, which calls for a different paper. My main purpose here was to show that this is, minimally, a difficult question. Put differently: to show that this part of the *Tractatus*, *contra* Anscombe, is not “most obviously wrong”.

References

Anscombe, G.E.M (1959) *An Introduction to Wittgenstein’s Tractatus*, 2nd edition, New York: Harper & Row.

Wittgenstein, Ludwig (1922) *Tractatus Logico-Philosophicus*, London: Kegan Paul.

Wittgenstein, Ludwig (1961) *Notebooks, 1914-1916*, New York: Basil Blackwood.

Friedlander, Eli (2018) "Logic, Ethics and Existence in Wittgenstein’s Tractatus", in: Reshef Agam Segal & Edmund Dain (eds.), *Wittgenstein’s Moral Thought*, London, New York: Routledge.

Stokhof, Martin (2002) *World and Life as One: Ethics and Ontology in Wittgenstein’s Early Thought*, Sanford: Stanford University Press.

There Is Still Work for the Fundamental

Tarun Thapar (New Delhi, India)

Abstract

Metaphysical Foundationalism is the view that chains of grounding must ultimately bottom out in something fundamental. Among the arguments that have been put forth for Foundationalism, the most developed argument is the externality argument in Bliss (2019) which states that there is an explanatory demand that any successful metaphysic should meet that the anti-foundationalist view like Infitism (that allows for infinite grounding descent) is not able to and as a result, exhibits an explanatory failure. Recently, the externality argument has been challenged by Oberle (2022) who claims that the foundationalist has failed to offer a convincing justification for their externality argument and this explanatory demand can be successfully met by the infitist. In this paper, I will try to resist this infitist's challenge to the externality argument. I will try to show that even if we grant the controversial assumptions to the infitist, then also, on at least some infitist chain, we don't get a complete explanation of the existence of non-fundamental entities.

Metaphysical Foundationalism is the view that chains of grounding must ultimately bottom out in some ungrounded fundamental entities i.e. every dependent entity must be fully grounded in some fundamental entity. Infitism allows for some chains of grounding to be infinitely descending without reaching a fundamental level. Here, grounding is understood as a relation of non-causal metaphysical determination that is tightly connected with metaphysical explanation. If an entity grounds another entity, the former also metaphysically explains the latter. I will talk in terms of entities to remain neutral on the debate about what kind of entities (facts, things, events etc.) can enter the grounding relation as the arguments discussed do not hang on this debate.

Although not unchallenged, foundationalism has been the orthodox view among metaphysicians. According to some philosophers like Bliss (2019, 2024), the most promising argument in favour of metaphysical foundationalism is "the externality argument" according to which, without fundamental entities, we cannot explain why there are any dependent entities whatsoever. The argument tries to secure a job for fundamental entities by creating an explanatory demand that can only be successfully met by the foundationalist. One crucial premise of the argument is the externality assumption: no dependent entity can explain why there are any dependent entities whatsoever. Recently, Oberle (2022) has tried to show that, on plausible

construal of the externality argument, the infinitist can meet this explanatory demand, by challenging the externality assumption.

In this paper, I develop a response that the foundationalist can offer against the infinitist argument offered by Oberle. Specifically, I will try to show that there is a crucial assumption- that there is a complete explanation of the existence of each member in the grounding chain- that the infinitist takes for granted and this assumption is false for at least some grounding chain. This in turn will result in a novel defense of the externality assumption.

1. Externality Argument- Work for the fundamental!

We find the most developed version of the externality argument in Bliss (2019: 369) -

1. There is an explanation for why there are any dependent entities whatsoever.
2. No dependent entity can explain why there are any dependent entities whatsoever.
3. Therefore, there must be something fundamental.

This argument works in two steps. In the first step, an explanatory demand is made from our metaphysician and in the second step, via the externality assumption (Premise 2), it is established that only a foundationalist can meet this demand successfully. Let me label the externality assumption as EA. The foundationalist claims that, according to EA, no dependent entity in the chain can meet this explanatory demand, therefore, to meet this demand, there must be something that lies outside the chain of dependent entities and that is the fundamental.

Since my aim in this paper is to analyze the infinitist argument developed by Oberle, the space doesn't permit me to discuss the justifications offered by Bliss for both the premises of the externality argument. However, before we get to Oberle's argument, we would at least need to get clear on what is our explanatory target in Premise 1.

Can our explanatory target be an existential fact: the fact that there are dependent facts? But, it is a widely accepted principle of logic of ground that the existential fact is explained by each of its instances. But then there is nothing left for the fundamental facts to explain regarding this existential fact. Perhaps, our explanatory target is a set, super-conjunction or a mereological fusion. Yet, again, on plausible assumptions regarding grounding, sets are grounded in their members, super-conjunctions in their conjuncts and mereological fusions in their parts. Hence, there doesn't seem to be anything missing to be explained if we understand our explanatory target this way.

Thus, given the implausibility of these above options, Oberle (p.989) suggests that we should understand the explanatory target as a plurality. Although Oberle talks in terms of facts, I keep talking in terms of entities to preserve consistency with Bliss' version of the externality argument. On the plurality construal, our target to explain is the existence of the plurality of all these dependent entities. As a result, what the foundationalist is seeking an explanation of is the existence of all the dependent entities taken plurally. On plurality construal, the new version of the externality assumption is: No dependent entity can explain the existence of the plurality of dependent entities. Let me label it as EA*. According to Oberle, a pertinent justification of EA* is that explanations that disobey EA* are incomplete in an important manner. Perhaps, the foundationalist's argument is that "infinite successive explanations without a fundamental ground can, at best, only offer us an incomplete explanation of the existence of the plurality of dependent facts" (p. 995). Accordingly, the foundationalist worries that if the chain does not end, everything in the plurality that needs a certain kind of explanation doesn't get it and thus, the plurality lacks a "complete explanation" of its existence whereby complete explanation we mean that nothing about an entity is left to be explained. To demand a complete explanation of an entity's existence is to demand an explanation of each and every aspect of its existence, it is to demand an explanation of all the facts related to its existence. Thus, on plurality construal, the foundationalist worry is that on infinitism, we lack a complete explanation of the existence of the plurality of dependent entities. I will return to the notion of a complete explanation in a later section. This will suffice our purpose for now.

2. Innitist Argument- No work for the fundamental!

Let us grant that the plurality construal is the most plausible construal of both the foundationalist's explanatory target and the corresponding externality assumption. Now, we are well placed to understand Oberle's criticism against the externality argument. Oberle develops an infinitist argument to challenge EA*. The infinitist argues that on the infinitist chain, the plurality gets a complete explanation of its existence.

Let's build his argument step by step. Firstly, Oberle (2022: 997) offers a ground-theoretic version of the Hume-Edwards Principle in the form of G-HEP-

G-HEP: The plurality of all dependent entities, dd , are fully grounded iff, for each fact x among dd , there are τ such that τ fully grounds x .

Note that G-HEP only holds if we allow for a distributive notion of plural ground. **Distributivity** entails that given two pluralities of entities, the individual members of each plurality are grounded in individual members of others.

I identify the next assumptions in Oberle's argument (p.997) as the following principles-

FG (Full Grounds): Every dependent entity in the infinitist's chain has full grounds.

FG-E (Full Grounding-Explanation): If Δ fully grounds E , Δ completely explains E 's existence.

Putting these together, we get the following infinitist argument against EA*-

1. If each dependent entity of the plurality of all dependent entities has a complete explanation of its existence, then the plurality has a complete explanation of its existence. [G-HEP + Distributivity + FG-E]
2. On infinitism, each dependent entity of the plurality has a complete explanation of its existence [FG + FG-E]
3. Therefore, the plurality of dependent entities has a complete explanation of its existence.
4. Therefore, there is no work for fundamental entities.

3. Denying Premise 2) of the Infinitist's Argument.

Let us grant the infinitist's argument as much as we can. Although, the foundationalist can resist each of the following principles, for the sake of argument, let us assume that G-HEP, Distributivity and FG holds. I will discuss FG-E in a later section. Now, even if we grant E-HEP, Distributivity and FG, the infinitist argument only works on the assumption that each member of the plurality receives a complete explanation of its existence. I believe that the main strength of the argument lies in Premise 2) of the infinitist argument and as a result, the foundationalist should try to deny Premise 2). To do that, the foundationalist needs to find at least one infinite dependence chain on which it's not the case that each member has a complete explanation of its existence. I claim that the foundationalist can indeed find such a regress.

Consider the relation of set membership. Now, it's a common view that sets are grounded in their members. We can use this relation to generate an infinite grounding regress if we have a set that has another set as its sole member, that set has another set as its sole member and so on... So, we get a grounding regress of the form: S1 is grounded in S2, S2 is grounded in S3, S3 is grounded in S4 and so on, ad infinitum, where S1, S2, S3, S4 and so on are all singleton sets.

In the context of our argument, we aim to determine the infinitist's assumption that each set has a complete explanation of its existence on this regress.

Allow me to build the foundationalist's argument step by step. Consider these three principles that I have adapted from Cameron (2022):

GG (Grounding Grounded): If E1 is grounded in E2, it is partially grounded by the fact that E2 exists.

Essence: If E1 is grounded in E2, it is part of the nature of E1 that E1 is grounded in E2.

EE (Existence-Essence): The complete explanation of E1's existence involves an explanation of the entirety of E1's nature.

Let's consider GG. We find versions of this principle endorsed by Bennett (2011) and de Rosset (2013). What's the rationale for GG? The argument, in its most general form, is that grounding unlike causation is an internal relation since it obtains in virtue of the nature of its relata. We only need the set and its members for the grounding relation to obtain. So, the grounding fact $\langle E1 \text{ is grounded in } E2 \rangle$ obtains either in virtue of (a) $\langle E1 \text{ exists} \rangle$ or (b) $\langle E2 \text{ exists} \rangle$ or (c) $\langle E1 \text{ exists} \rangle$ and $\langle E2 \text{ exists} \rangle$. Now, either option (b) or (c) would suffice to grant GG since we need our grounding fact to be just partially grounded by $\langle E2 \text{ exists} \rangle$. So, we only need to eliminate option (a). Option (a) says that members ground their set solely in virtue of the fact that sets exist. Now, this seems quite implausible that to make sets dependent on their members God would have needed to create just sets. It is a general view that it's not the sets but rather the members that are ontologically prior to sets. As a result, option (a) seems very implausible. Thus, if GG holds as a general principle of grounding, it holds for the relation of set-membership as well.

Does Essence hold for the set-membership relation? It says that it is part of the nature of the set that it is grounded in its members. I think Essence is extremely plausible if we take sets to have their members essentially. If we take the Finean (1995) way to characterize grounding in terms of a non-modal notion of essence: $E1$ is grounded in $E2$ iff $E2$ is a constituent of an essential property of $E1$, then, the essence of the grounded entity (set) involves the entity that grounds it (members). As Cameron (2022: 99) puts it—"The set can't exist—it cannot be that very set—without having each of the Xs as a member and, as a result, being ontologically dependent on each of the Xs ". It is in the nature of the set that it has the members it has and this makes it the case that it is in the nature of the set that its members ground it. Thus, it is in the nature of a singleton not only that it has the member it has but also that it is grounded in its member.

What about EE? The idea behind this principle is that, under very plausible assumptions, the notion of existence and the notion of essence have a very tight connection. If one needs to completely explain the existence of an entity- why that entity exists rather than some other entity- then one would have to explain why something with that entity's nature exists. I will come back to this principle in a later discussion.

Now, we are in the position to check whether each set has a complete explanation of its existence on our infinitist grounding regress.

Let's begin with S1 and try to give a complete explanation of its existence. We would need to explain all the facts about S1 such that nothing is left to be explained about S1's existence. Now, given EE, to completely explain the dependent entity S1's existence we would need to explain S1's nature as well. So, what explains why S1 exists and the nature it has? It's the fact that S2 exists and the fact that S1 is grounded in S2. Now, given Essence, the fact that S1 is grounded in S2 is part of the nature of S1 which we need to explain in the first place since we are also looking for the explanation of S1's nature. Thus, the fact that S1 is grounded in S2 is part of what we want to explain and what is being used to explain. So, we would need to explain it. But, given GG, this fact's explanation involves explaining E2's existence. Thus, for the success of the complete explanation of S1's existence, we need to successfully explain the existence of S2. And repeating the same reasoning, for the success of the complete explanation of S2's existence, we need to successfully explain the existence of S3 and so on ad infinitum. As a result, our original explanation for why S1 exists and has the nature it has is hostage to explaining the existence/nature of S2 and so on...the explanation is never successful. The success of the explanation is always postponed, never established. We never manage to successfully explain the existence and nature of S1. Thus, if there is no last link in the chain, the success of our explanation will be indefinitely postponed and our explanation will never be complete. It is only when there is a last link in the chain, that our explanation will be complete and thus, successful. And the last link is at the fundamental level. Hence, contra Infinitist, we don't get a complete explanation of the existence of each member of the plurality. As a result, Premise 2) in the infinitist argument is false.

4. An Objection & Reply

Here, I can think of at least one peculiar sort of objection that the infinitist can raise against the argument that I have given above. The objection concerns FG-E. It is a universally acceptable grounding principle that members fully ground their sets. So, if we have a singleton {Socrates}, the individual Socrates fully grounds {Socrates}. Hence, the objection is that if a member fully grounds its

singleton then the singleton's existence is completely explained by its member's existence. Nothing is left to be explained about the dependent entity if it is fully grounded in another dependent entity. So, in the set-membership regress, we get a complete explanation of each set's existence in terms of its full grounds.

I think the foundationalist can respond to this objection by claiming that it is an error to assume that full grounds generate complete explanations. Full grounding explanations at the derivative level are not always complete explanations because it can be the case that a dependent entity on the infinitist chain receiving full grounds nevertheless lacks a complete explanation of its existence. As I have said earlier, to demand a complete explanation of an entity's existence is to demand an explanation of each and every aspect of its existence. We are not just asking why the entity exists but also asking why that entity exists instead of some other entity or instead of not existing. And this includes accounting for its essence as there can't be a set with that nature without it being the very set. EE gives the foundationalist a way to link the explanation of an entity's existence with the explanation of its essence. If one aims to "fully explain why E1 exists—why that particular thing exists, rather than something else—then that is *inter alia* to explain why something with E1's nature exists" Cameron (2022: 102). Thus, in our case, though S2 fully grounds S1, it doesn't fully explain S1's existence for the reason that it doesn't explain why S1 has the nature it has. It is for this reason that FG-E doesn't hold. Since, our concern is with the complete explanation, thus, in our case, the foundationalist's demand of explaining S1's existence includes explaining S1's nature and given Essence, it involves explaining why S1 is grounded in S2 and this involves, given GG, explaining why the particular set S2 exists...and so on *ad infinitum*. And, if there is no last link in this chain our explanation will never be complete. It is only when there is a last link, some fundamental entity, that our explanatory demand would be met. It is for this reason that every dependent entity in this chain needs to be fully grounded in the fundamental entities.

5. Conclusion

In this paper, I have tried to develop a strategy for the foundationalist to defend the externality argument against the infinitist's challenge. And, in the process, we have got a novel defence of the plurality version of the externality assumption based on some principles related to grounding and explanation that haven't received significant attention in the literature and I hope that I have at least been able to motivate these principles plausibly. However, this has been only a partial defence of the externality argument. The complete defence involves justifying the explanatory demand in the first place. And that in turn involves motivating some version of the Principle of Sufficient Reason. And that is no small task but there is still work for the fundamental and the foundationalist should get to the task.

References

- Bennett, Karen (2011) "By Our Bootstraps," *Philosophical Perspectives* 25, 27–41.
- Bliss, Ricki Leigh (2019) "What Work the Fundamental?," *Erkenntnis* 82(2), 359–79.
- Bliss, Ricki Leigh (2024) *Grounding, Fundamentality and Ultimate Explanations*. Cambridge: Cambridge University Press
- Cameron, Ross P. (2022) *Chains of Being: Infinite Regress, Circularity, and Metaphysical Explanation*. Oxford: Oxford University Press.
- deRosset, Louis (2013) "Grounding Explanations", *Philosophers' Imprint* 13, 1–26.
- Fine, Kit (1995) "Ontological Dependence", *Proceedings of the Aristotelian Society* 95, 269–90.
- Oberle, Thomas (2022) "No Work for Fundamental Facts", *Philosophical Quarterly* 73 (4): 983-1003.

Agency in Virtual Environments

Sabine Thuermel (Munich, Germany)

Abstract

The production and reception of fiction can be seen as an anthropological constant. However, the specific characteristics of works of imagination and their reception are specific to both (cultural) epochs and media. Virtual environments have been widely used for several years, and AI has recently become increasingly supportive. Therefore, a short presentation of virtual environments in general is given followed by an introduction to the latest development in this field: generative AI-empowered environments. This lays the foundation to take a closer look at how virtuality emerges from computer mediation and how generative agency differs from human agency. It is demonstrated that the expectations of generative agents and their social behavior should remain modest: our living environment, into which humans are born without dependence on training data, offers completely different development possibilities than generative agents possess based on machine token operations and large language models. The paper concludes with an outlook on how Baudrillard's ideas on simulation and simulacra can be adapted to generative AI-empowered environments. It is shown that the fictional-virtual-digital triad is particularly evident in the generative spaces: digitality lays the foundation for experiencing virtual spaces and for interacting with (generative) software agents and avatars that are controlled by human users. This allows new imaginary worlds and novel fictions to emerge.

1. Introduction to computer-mediated virtual environments in general

Virtual environments are "interactive mirrors of dynamized symbol worlds", to use an image developed for computers in general by Sybille Krämer in (Krämer 2011, p.314). They are based on the possibilities of the digitalized representation of symbol structures, their (graphic) representation, and computer-based modifications. They offer a "real-time interaction space" (Lattemann 2008). These spaces are characterized by the fact that the actors - users and programmers - have a share in their design. Moreover, these spaces are persistent. Therefore, they continue to exist even if a user leaves them. In contrast to the living world that surrounds us materially, they are interactive "computer-based simulated environments" (Søraker 2010, p. 30ff) whose dynamic appearance is represented online. Their multitude of applications covers the entire spectrum of the term "simulation": from everyday use in the sense of "deceptive" - as in gameworlds such as World of Warcraft® - to social networks with their diverse possibilities for self-expression and communication, to scientific knowledge-generating numerical simulations of mathematical models.

Virtual worlds can be realistic, abstract, or purely imaginary. The processes that take place there can (seemingly) resemble those of our living world or be implemented according to their rules governing specific imaginary worlds. Users can create their own avatars, which they use to navigate and act in the world. The players can help shape the environment and interact with their communication partners - be they avatars or pure software agents: Virtual worlds allow human users - usually via avatars (i.e. graphical representations) - to engage in role-playing games, cooperate in games according to rules, and generally collaborate and pursue specific goals, also in the latest virtual business environments. Users can be physically distributed around the world and yet still be (tele)present in the virtual world. They only show themselves there via their virtual person.

Pure software agents, i.e. those that are not controlled online by humans, are used in various forms: as assistants/software-based tools and as interaction partners in online games, but also in professional environments and even in legally binding contexts such as online auctions. Only software agents can be found in artificial life systems or simulations of technical or natural processes.

Virtual environments differ from classical fictional worlds expressed via other media in that their specific form of virtuality results from being computer-mediated. Moreover, the mediation offers their designers to create spaces where human users may experience artificial agency. These two aspects will be outlined in more detail in the next two sections using as examples the latest development in the long evolution of software agents, namely generative agents. To familiarize all readers with generative agents, a short introduction to agency based on Generative AI is given in the next paragraph. After that, the potentiality and actuality of virtual environments, especially AI-empowered environments are described. Artificial Agency in virtual environments is introduced and exemplified again by generative agents. This exemplary focus is chosen because the capabilities of generative agents are based on text and image thesauri on the Internet, i.e. our digital environment.

2. Introduction to Generative AI-empowered Environments

Generative AI is the latest development in AI and its currently most prominent area, namely machine learning (ML). The most important innovations of this

ML field lie in the use of large language models and the mechanism of self-attention (Vaswani et al., 2017): Large language models (LLMs) can process large amounts of unstructured text and describe the relationships between words or parts of words by using the mechanism of self-attention to efficiently capture complex relationships between different parts of the input sequence. Thus, Generative AI can learn patterns, e.g. speech patterns, at extremely high speed, depending on access to training data, e.g. texts. It is therefore pre-trained. Depending on the input (usually in natural language), stochastic algorithms generate new digital content that is "as similar as possible" to the input. Due to this non-deterministic approach, every generative AI system exhibits weak emergence: Each time it is used with the same input, it can produce a different output that cannot be predicted, but only emerges during runtime. Looking at these systems, it is therefore not surprising that Floridi chose the following perspective on AI, namely to "conceptualize AI as a growing resource of interactive, autonomous, and often self-learning agency (in the ML sense) to deal with tasks that would otherwise require human intelligence and intervention to be performed successfully" (Floridi 2023, chapter 2). Even today Generative AI environments may support or even automate the decision-making and work processes of knowledge workers of all kinds and creative professionals. However, it still has to be proven whether AI truly provides a "reservoir of smart agency on tap" (Floridi 2023, chapter 8).

A special case of AI-empowered agency is software agents exploiting an artificial agency approach to model virtual agency. Even today LLM-based agents may be found in a wide variety of application fields. The currently realized scenarios encompass a range of contexts, including a single software agent executing tasks, collaborative and competitive systems of software agents, and software agents assisting humans (Xi et al. 2023). If one considers the fact that current AI systems are engineered by humans, Floridi's dictum that AI is a "divorce (not a marriage) between agency and intelligence" (Floridi 2023) rings true for all current systems.

3. Potentiality and Actuality of Generative Agents in Virtual Environments

In technology-based contexts, virtuality emerges from computer mediation. The potentiality inherent in computational creations is actualized during

execution. Using software engineering methods, each of these artifacts is designed to run on a computer, i.e. an instance of a universal Turing machine. It is therefore also an abstract object (type), which explains its proximity to pre-technical virtuality objects such as Leibnizian monads or mathematical objects. Every computer-mediated artifact is intangible, i.e. it exists in time but not in space. The computer is used as a medium that allows the sensory experience and manipulation of what is experienced within the limits set by the environment. As a sign-processing machine, the computer enables the (arbitrary) circulation of signs in computer systems (without media breaks). Thus, although the computer in von Neumann's architecture is an instance of the universal machine, it is not a universal media machine, but a unique media machine. All these virtual environments function due to the performativity of the underlying system, a computational artifact. The potential of such systems and their potential sociality becomes actual in a concrete instantiation. If Hubig's philosophy of technology as a reflection on the mediality of all technology (2006) is applied to computer-mediated systems, the specification of a virtual environment corresponds to Hubig's "possibility space of the realization of possible purposes". An instantiation at runtime corresponds to Hubig's "reality space of possible purposes". Regarding its effect on the physical environment, an experimental environment possesses a virtual actuality and a (cyber-physical) system that controls processes in the physical world has a real actuality. A case in point is the interactive simulacrum of human behavior by generative agents realized by Park et al. in 2023: it is demonstrated in a sandbox of 25 LLM-based agents that "starting with only a single user-specified notion that one agent wants to throw a Valentine's Day party, the agents autonomously spread invitations to the party over the next two days, make new acquaintances, ask each other out on dates to the party, and coordinate to show up for the party together at the right time". They display their sociality when organizing a party in "Smallville" all by themselves in a simulacrum of virtual actuality. Such simulacra possess a real actuality where humans interact with these generative agents or when processes in the physical worlds are controlled by generative agents e.g. when buying online for their prompt masters. The degrees of freedom of generative agents, their social behavior, and their interactions with their human partners are determined by the environments in which they take place. Therefore, the expectations of generative agents and their social behavior should remain

modest: our living environment, into which humans are born without dependence on training data, offers completely different development possibilities than generative agents possess based on machine token operations and large language models.

4. Simulacra and Simulation

Baudrillard's work on "Simulacra and Simulation" (1994) immediately comes to mind when considering the diverse application scenarios of generative agents. The four stages "faithful copies of reality", "perversions of reality", "the absence of profound reality" and "pure simulacra" coexist not only in mass media, as Baudrillard stated, but also in Generative AI-empowered environments: The bionic and socionic approaches represent more or less "faithful copies of reality" e.g. in the electronic auctioning systems for matching supply and demand of scarce resources. Some human players use generative AI to provide "perversions of reality" e.g. to defame other humans. In purely fictional environments "the absence of a profound reality" and "pure simulacra" may be found making the sociality between generative agents in these contexts a simulation. However, these environments go beyond what characterizes mass media: New hybrid environments come into being, where humans interact with virtual agents, e.g. with virtual influencers or their virtual doppelgänger in a simulacrum or with bots providing a version of "smart agency on tap" (Floridi 2023, chapter 8). They offer real-time interaction spaces where humans may explore the sociality and the digital content provided in these spaces. In contrast to the late 1990s, such virtual environments are pervasive in our current culture. Natural language interaction has replaced scripted interactions. Baudrillard's conviction that there is a "transition from signs which dissimulate something to signs which dissimulate that there is nothing" is also true for any sign of social intelligence displayed by generative agents if you take "nothing" to mean acting without truly comprehending the meaning of sociality for us humans. Consequently, adult humans must be aware as children always are that pretend play is quite different from authentic human sociality.

5. Conclusions

The production and reception of fiction can be seen as an anthropological constant. However, the specific characteristics of works of imagination and their reception are specific to both (cultural) epochs and media: Virtual worlds, i.e. computer-based simulated environments with which both human users and (pure) software agents interact, can be seen as exemplary for this thesis. These software systems not only provide a participative medium of expression but also, for the first time, a medium of experience in which human users interact with each other and with pure software agents. The experiential spaces created in virtual worlds offer room for cognitive and communicative action as well as symbolic interaction, i.e. for skills that are accorded special importance in our era. The fictional-virtual-digital triad is particularly evident in the generative spaces: digitality lays the foundation for experiencing virtual spaces and for interacting with (generative) software agents and avatars that are controlled by human users. This allows new imaginary worlds and novel fictions to emerge.

References

- Baudrillard, Jean (1994) *Simulacra & Simulation*, Michigan: University of Michigan Press.
- Floridi, Luciano (2023) *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*, Oxford: Oxford University Press.
- Hubig, Christoph (2006) *Die Kunst des Möglichen I – Technikphilosophie als Reflexion der Medialität*, Bielefeld: Transcript Verlag.
- Park, Joon Sung, et al. (2023) "Generative Agents: Interactive Simulacra of Human Behavior", *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1-22.
- Vaswani, Ashish et al (2017) "Attention is All You Need", *Computation and Language (cs.CL); Machine Learning*, arXiv:1706.03762
- Xi, Zhiheng, et al. (2023) "The Rise and Potential of Large Language Model-based Agents: A Survey". *arXiv preprint arXiv:2309.0786*

Modal Realism and the PSR

Tarik Tijanovic (Santa Cruz, USA)

Abstract

Peter Van Inwagen argues that The Principle of Sufficient Reason (PSR) either leads to contradiction or it leads to necessitarianism. Although I agree with Van Inwagen that the relationship between the PSR and necessitarianism is close, I argue that the PSR is compatible with innocent versions of necessitarianism. In this project my main argument is that modal realism can account for the PSR and integrate it within an innocent version of necessitarianism. My main claim will be that each fact in each possible world is grounded in the principle of plenitude. Also, while each fact is grounded in plenitude, it avoids rigid necessitarianism because (i) actuality is not captured by the PSR, (ii) things could be otherwise because there are other worlds, and due to the principle of recombination (iii) cardinality of possible worlds is maximized and (iv) there are no necessary connections between distinct individuals. As a result, the relationship between modal realism and the PSR is symbiotic. Through modal realism, the PSR also results as: having its scope extended to the realm of all possibilities rather than being confined to mere actuality; it solves indeterminacy problems like Leibniz's problem of space or the hard problem of consciousness; it accounts for facts in lawless worlds etc. On the other hand, through the PSR, modal realism gets one more benefit in its inventory because it accounts for the PSR. Thus, modal realism still continues on Lewis' paradoxical path of proving to be highly pragmatic but counterintuitive at the same time. Ultimately, even if we reject modal realism, we can use it as a model for future defenses of the PSR against rigid necessitarianism.

The principle of sufficient reason (the PSR) states that every fact has a sufficient reason. Van Inwagen presented a counterargument to the PSR which was for a long time considered a refutation of the PSR. Here is the presentation of Van Inwagen's argument by McDaniel:

Assume that there is at least one contingent truth. Call the conjunction of all contingent truths 'C'. C is contingent, since any true conjunction with at least one contingent conjunct is itself contingent. So C has a sufficient reason; call it 'R'. R is either contingent or necessary. If R is contingent, then R is a conjunct of C. But a contingent conjunct of a conjunction cannot explain that conjunction. If R is necessary, then C is necessary, since sufficient reasons entail what they explain. (2016:1)

McDaniel continues:

One can derive a formal contradiction from what was stated in the previous paragraph. Something has to go. van Inwagen rejects PSR and suggests that we do the same." (2016: 1)

According to Van Inwagen, the PSR either leads to contradiction or to full-blown necessitarianism. Many tried to defend the PSR by appealing to grounding relations “instead of appealing to an underspecified notion of explanation” (McDaniel, 2016:2). One solution is to say that a contingent conjunction is grounded in individual conjuncts. Therefore, the PSR is accommodated without necessitarianism. However, Van Inwagen easily avoids the problem by saying that the series of conjuncts need grounding rather than a single conjunction. In that case, the series of individual conjuncts would be grounded in themselves. That would lead to the violation of irreflexivity which is one of the hallmarks of grounding.

In this paper I defend the PSR against Van Inwagen’s argument by grounding it in an abstract principle. In order to do that I will use the modal realist framework. Here I briefly introduce the framework. According to modal realism, all possible worlds exist. The crucial principle is the principle of plenitude which says that all possible worlds must exist, not some or many. The principle of recombination underlies the principle of plenitude and roughly states that other possible worlds are recombinations of the actual world. Also, since possible worlds are causally and spatiotemporally isolated from each other, actuality is an indexical. For members in our world, our world is actual while others merely possible. For members of other worlds, their worlds are actual while ours is merely possible for them.

With such a framework in place, I go on and defend the PSR. The way I will defend the PSR is by agreeing with Van Inwagen that the relationship between the PSR and necessitarianism is close, but I will disagree that necessitarianism is an all-or-none phenomenon. In fact, I will show that there are innocent versions of necessitarianism that can account for the PSR.

First of all, I will say that the sufficient reason for each fact in each world is the principle of plenitude. If we ask: why does X exist? The answer is: because it is grounded in the principle plenitude (it must exist in some world). So far we have rigid necessitarianism—each possible fact necessarily has to be placed in some world. If one finds such an explanation boring, we can opt for Lewis’ counterfactual theory of causation to substantiate it. Namely, we can explain the reason for individual’s existence not directly in plenitude but in relation to other individuals. For instance, we can say that X exists because X is caused by

Y or X is grounded in Y. And further, for X to be caused or grounded in Y is for there to be another world in which X is caused or grounded in Y which is closer than a world in which X is not caused or grounded in Y. One can choose an alternative relation, for instance: X exists because there is a possible causal or grounding relation between X and Y. There is no problem with unpacking plenitude in terms of causation given that there is the plenitude of all possible causal or grounding relations.

This is how each existing fact gets its sufficient reason. Here is how necessitarianism gets mitigated by modal realism. First, actuality is not captured by the PSR. Since actuality is indexical, there can be no reason why certain worlds are actual while some others are not. While we should not be happy with any restriction of the PSR in general, the restriction of the PSR from actuality is not problematic. Each existing fact in each world is still covered by the PSR.

Second, modal realism allows for things to be otherwise. The mere fact that there are other possible worlds shows that the modal realist universe is not rigidly necessitarian. For instance, for Spinoza there are no other possible worlds, the only possible world is the actual world which is the manifestation of God's essence. For Leibniz there are possible worlds but they are not actualizable because the actualizer—God—is constrained by the principle of the best and needs to create the best out of all possible worlds. For thinkers like Hume, there are abstract possible worlds that could be actualizable but Hume with his no-necessary connection view dissolves the PSR. He allows for brute facts for instance. The point is that when we talk about other possible worlds, they both need to be “actualizable” and covered by the PSR, for necessitarianism to be innocent. So far only modal realism allows for that.

Third, if we rely on the principle of recombination we maximize the cardinality of possible worlds. The principle of recombination allows for lawless worlds or even for Leibnizian “harmonious” worlds without causal relations. The fact that the number of worlds in logical space is maximized says something about the actual world—that it is not rigidly necessitated.

Fourth, the principle of recombination is the Humean principle which says that there are no necessary connections between distinct individuals. Living in

a Humean universe further minimizes necessitarianism. The reason why I coexist with a horse is simply because we are distinct; in some other worlds I exist without a horse.

Modal realism doesn't only benefit the PSR because it grounds each fact in the principle of plenitude while minimizing necessitarianism. It provides more benefits. First, it accounts for indeterminacy problems which are at the core of the PSR. For instance, consider Leibniz's problem of indeterminacy of spatial rotations. According to Leibniz, God couldn't decide which rotations of space in relation to objects to choose. Since he couldn't make a preference for one rotation over another, God didn't create space at all. The point is that since God acts according to the PSR, God had to avoid the creation of space. As we see, the problem is serious to the point that Leibniz had to avoid the claim that our world is in space.

Modal realism has the solution for this problem and all other indeterminacy problems. For each indeterminacy, there is a plenitude of worlds in which that indeterminacy is resolved across possible worlds. For instance, if we ask why the actual world has one rotation over another, we simply say that it has to have because of plenitude. That is, there is the plenitude of rotations of worlds in logical space and each world must take up one rotation. Of course, to rehearse details we would need counterpart relations between spacetime points or we would even need to posit real coordinate systems to be able to compare rotations across possible worlds.

There are similar indeterminacy problems that can also be tackled by modal realism. For instance, one indeterminacy problem is the hard problem of consciousness. Once neuroscientists discover a neural correlate of consciousness, we can still ask why this neural correlate rather than some other one? Again, with modal realism in place we can spread the correlates across possible worlds and say that in each world, consciousness had to have a possible correlate, or that in some worlds there are zombies (consciousness comes apart from any neural correlate).

Besides indeterminacy problems, modal realism tackles brute facts which are known to be incompatible with the PSR. For instance, within the framework of modal realism a world in which gravity suddenly fails is not a brute fact but a

fact that can be explained in plenitude. Since such a world is possible, it has to exist according to the principle. The upshot is that there are no brute facts in the modal realist universe. Furthermore, the PSR's scope is not only extended into the realm of possibilia, but to the realm of all possibilia—it is necessary.

The upshot is the following. We can still ground existing facts without inducing rigid necessitarianism. With modal realism we achieve it by avoiding capturing actuality under the PSR, by allowing other possible worlds, by using the principle of recombination and by maximizing cardinality of possible worlds.

While writing this paper I could not suggest that my main motivation is to provide the account for the PSR given that one needs to “swallow” modal realism first, which barely anyone wants to. If one doesn't want to accept modal realism in order to accept the PSR, one can still benefit from this inquiry. First, one can better appreciate problems that the PSR and one can learn more about modal realism. For instance, one can see this project as the continuation of Lewis' project where modal realism always proves to be pragmatic—it provides theoretical unity and economy—while never being accepted because the price is too high. Even more strongly, one can say that modal realist account of the PSR is at the top of its benefits.

More importantly, one can benefit from this inquiry because one can see it as a model for future defenses of the PSR. I showed that we can have the PSR and necessitarianism without having rigid necessitarianism. Although most of the ways of minimizing necessitarianism presented here rely on a modal realist framework, there is something that all future frameworks can have in common—they can say that things could be otherwise in an innocent necessitarian universe. I claim that even without a modal realist framework we could have other possible worlds that could be actualizable. For instance, let's take Al-Ghazali's voluntarist God as an example. Such God creates the actual world based on free will rather than by constraint of his own benevolence as is the case with Leibniz's God. If we live in a world created by God's free will, then everything in our world has a sufficient reason and still, things could be otherwise. God could choose to actualize the actual world in some other way. He could even annihilate this world and create time anew. We have the PSR without necessitarianism.

Another way of accounting for the PSR without necessitarianism is opting for Rescher's de-theologized version of the principle of the best (Rescher, 2000). According to Rescher, the reason for everything in existence is because it is the best possible way, while the world is not actualized neither by God, nor by any kind of agent. Instead, the actual world exists because "less than the best scenarios are eliminated". Rescher thinks that human freedom must be included in such a world which again, avoids necessitarianism.

Also, while Rescher's and mine account use grounding relations, Al-Ghazali's example uses causal relations. Different kinds of relations can play a role in accounting for the PSR. Furthermore, as I showed (and as we can see with Rescher's account), I grounded each fact in the principle of plenitude while Rescher explained each fact in the principle of the best.

References

- al-Ghazālī (1927) *Tahāfut al-Falāsifah*, M. Bouyges (ed.), Beirut: Imprimerie Catholique.
- al-Ghazālī (1962) [AIFI], *Al-Iqtisād fī l-i'tiqād*, I.A. Cubukcu and H. Atay (eds.), Ankara: Nur Matbaasi.
- al-Ghazālī (2000) [IP] *The Incoherence of the Philosophers*, M. Marmura (trans.), Provo: Brigham Young University Press.
- Amijee, Fatema (2021) "Principle of Sufficient Reason", in: D. Jalobeanu and T. Wolfe (eds.), *Encyclopedia of Early Modern Philosophy and Science*, Springer.
- Amijee, Fatema (2022) "Inquiry and Metaphysical Rationalism", *Australasian Journal of Philosophy*.
- Chalmers, D.J. (2002) "Does Conceivability entail Possibility?", in: Tamar Gendler & John Hawthorne (eds.) *Conceivability and Possibility*. New York: Oxford University Press. pp. 145--200.
- Chalmers, D. J. (1996) *The conscious mind: In search of a fundamental theory*, Oxford University Press.
- Rocca, Michael Della (2010) "PSR", *Philosophers' Imprint* 10.
- Du Châtelet, Emilie (2009) *Selected Philosophical and Scientific Writings*, Edited with an introduction by Judith P. Zinsser. Chicago: University of Chiicago Press.
- Horwich, Paul (1982) *Probability and Evidence*, Cambridge: Cambridge University Press.
- Hume, David (1978), *Treatise of Human Nature*, 2nd edition, Oxford: Clarendon Press.

Kant, Immanuel (1998) *Critique of Pure Reason*, Cambridge ; New York, Cambridge University Press.

Kripke, S. (1980) *Naming and Necessity*, Oxford: Basil Blackwell.

Leibniz, Gottfried Wilhelm; Clarke, Samuel & Ariew, Roger (2000) *Leibniz and Clarke: Correspondence*. Indianapolis: Hackett Publishing Company, Edited by Samuel Clarke & Roger Ariew.

Leibniz, G.W. *Sämtliche Schriften und Briefe*, Deutsche Akademie der Wissenschaften, multiple volumes in 7 series, Berlin: Akademie Verlag; abbreviated, cited by series, volume, and page.

Lewis, David K. (1986) *On the Plurality of Worlds*, Malden, Mass.: Wiley-Blackwell.

Lewis, David K. (1968) "Counterpart Theory and Quantified Modal Logic", *The Journal of Philosophy*.

Lewis, David K. (1973) *Counterfactuals*, Oxford: Blackwell.

McDaniel, Kris (2019) "The principle of sufficient reason and necessitarianism", *Analysis* 79 (2), 230-236.

Van Inwagen, Peter (1983). *An Essay on Free Will*, New York: Oxford University Press.

Pruss, Alexander R. (2006) *The Principle of Sufficient Reason: A Reassessment*, Cambridge: Cambridge University Press.

Rescher, Nicholas (2000) *Nature and Understanding: The Metaphysics and Method of Science*. Oxford, GB: Oxford University Press UK.

Spinoza, Baruch (1677) *Ethics: Masonic Edition*, Hackett.

Wolff, Christian (1720) *Rational Thoughts on God, the World and the Soul of Human Beings*, in: Eric Watkins (ed.), *Kant's Critique of Pure Reason: Background Source Materials*, Cambridge: Cambridge University Press, 2009, 7–53.

Engineering Concepts: Of Idling and Hot-Running Wheels

Jasmin Trächtler (Dortmund, Germany)

Abstract

Wittgenstein famously wrote that “philosophy may in no way interfere with the actual use of language” but can “only describe it” in order to ‘therapeutically’ treat and resolve philosophical-conceptual problems (PI 2009: 124, 133). Such a view is diametrically opposed to the goals of conceptual engineering: Rather than examining the meaning or use of concepts in ordinary language, conceptual engineering is concerned with the normative question of what concept ought to mean. Ameliorative conceptual engineering has also been used in social and feminist philosophy: It is argued that to achieve the political goals of social justice, it is also necessary to revise concepts, which in their ordinary meanings partly contain or reinforce injustice (e.g. “woman”, “black”, “misogyny”), i.e., feminist conceptual engineering interferes with the actual use of language.

In this paper, I will critically examine the relationship between Wittgenstein’s descriptive philosophy and some ameliorative projects in conceptual engineering. May we as philosophers really in no way interfere with the use of language? Not even when our concepts do not seem to fit our lives (anymore)? Wouldn’t we as philosophers and conceptual workers be particularly suited to adapt unsuitable concepts to our life needs? To answer these questions, I will first discuss some aspects of Wittgenstein’s metaphilosophy before taking a closer look at the aspirations behind some approaches to feminist conceptual engineering. I will conclude with a résumé of the relationship between Wittgenstein and feminist conceptual engineering, and of philosophy’s role in conceptual change in the service of social justice.

1.

Wittgenstein famously wrote that “philosophy may in no way interfere with the actual use of language” but can “only describe it” in order to ‘therapeutically’ treat and resolve philosophical-conceptual problems (cf. PI 2009: 124, 133). Thus, according to Wittgenstein, it is not the task of philosophy, like the (natural) sciences, to *explain* or *justify* certain phenomena – both linguistic and non-linguistic – nor to discover their ‘hidden’ causes or to set up hypotheses or theories (cf. PI 2009: 109). Philosophical problems are not of an empirical, but of a conceptual nature and arise when we philosophers allow ourselves to be misled by language. However, this doesn’t mean that “the most we can hope for” is *just* “the elimination of our traditional concerns”, after which “we will not be left with any [...] new understanding” (Horwich 2012: 20, 6). According to Wittgenstein’s conception of philosophy, the traditional (metaphysical) problems would indeed be dissolved (cf. BT 2005: 421), but for the sake of a new perspective on them, a new understanding.

The new perspective or new understanding of the object of philosophical investigation is not achieved by means of (hypo-) theses or by setting up new idols, but by means of a new organisation:

A philosophical question is similar to one about the constitution of a particular society. – And it's as if a group of people came together without clearly written rules, but with a need for them [...]. Thus they in fact view one of their own as president, but he doesn't sit at the head of the table and has no distinguishing marks, and that makes negotiations difficult. That is why we come along and create a clear order: we seat the president at a clearly identifiable spot, seat his secretary next to him at a little table of his own [...] etc. (BT 2005: 415)

Philosophy should thus create an order in our concepts so that their rules of use emerge clearly – but it doesn't bring to light any new or hidden facts, rather, as Wittgenstein famously stated: “[t]he aspects of things that are most important for us are hidden because of their simplicity and familiarity” – one is unable to notice them precisely because they are “always before one's eyes” (PI 2009: 129). E.g., he reminds us that the phenomena we in everyday life refer to by concepts such as “game”, do indeed have certain similarities, family resemblances, but they do not have *one* common feature, one necessary and sufficient criterion, based on which we label them “game”. *This* is one of the important aspects, which is so simple and ordinary that we don't notice it and which we philosophers – I dare say – have learnt or understood through Wittgenstein's philosophy. And seeing this requires no scientific explanation, justification or theory, but a description of the use of language.

However, providing such a description is easier said than done, because the description must “not interfere with the actual use of language”, while it should emphasise the “important aspect” that is misunderstood in philosophy, so the description cannot simply be a mere copy of language use.

What we call “*descriptions*” are instruments for particular uses. Think of a machine-drawing, a cross-section, an elevation with measurements, which an engineer has before him. Thinking of a description as a word-picture of the

facts has something misleading about it: one tends to think only of such pictures as hang on our walls, which seem simply to depict how a thing looks, what it is like. (These pictures are, as it were, idle.) (PI 2009: 291)

A description in the sense of such a “word-picture” would be just as useless as trying to draw someone’s attention to the “rabbit” aspect in the duck-rabbit-figure by making an exact copy of the figure. Although many lines of actual language use must be *traced* in the description, to emphasise the philosophically important aspects of language use, we must also *overdraw*, *rearrange* and even *invent* certain features or *supplement* them with ‘auxiliary lines’ so that we can clearly see the essential features of the *rabbit* and understand the figure (cf. RPP II 1998: 47). In this sense, the aim of philosophy is to offer an orientation aid with which we can find our way through the conceptual errors and confusions, i.e. philosophical activity is similar to “travelling in a landscape with a view to constructing a map” (RPP I 1998: 303). And we need different maps for different purposes and areas, and we also need new maps when the conceptual landscape changes. But it is true that Wittgenstein does not interfere with the landscape itself, i.e. ordinary language, with his grammatical investigations. His aim is to treat the philosophical ailment of language by pointing out and solving “the injustices of philosophy” (BT 2005: 420); but not to regulate or reform the ordinary use of language (cf. PI 2009: 130, 132).

Thus, Wittgenstein’s descriptive philosophy of ordinary language seems to be irreconcilably opposed to the explicit normative endeavours of so-called “conceptual engineering”.

2.

“Conceptual engineering” is a more recent term for the much older endeavour of changing and improving our ordinary concepts for various purposes. Conceptual engineering can be understood as “the process of designing, implementing, and evaluating concepts” by revising, fixing, abandoning and creating concepts (Chalmers 2020: 1-7). However, there is no unified understanding of what “conceptual engineering” includes and what it doesn’t, or whether philosophy doesn’t always include *some* form of conceptual engineering (cf. Cappelen & Plunkett 2020, 14).

As an early “key example” of conceptual engineering in social philosophy, Haslanger’s seminal paper on Gender and Race (2000/2012) is often mentioned where she aims at developing definitions of the concepts “gender” and “race” that aren’t simply an elucidation of the everyday use of these terms, but an “account of gender and race that will be effective tools in the fight against injustice” (ibid.: 226). Contrary to the everyday use of, e.g. “gender” in the 1990s as a polite way to talk about the sexes referring to conventional primary and secondary sexual characteristics, Haslanger wants to revise this use by the following definition of “woman”:

S functions as a woman iff_{df} (i) S is “observed or imagined to have certain bodily features presumed to be evidence of a female’s biological role in reproduction”, which (ii) marks S as “someone who ought to occupy certain kinds of social position that are in fact subordinate” and (iii) the fact that S satisfies (i) and (ii) plays a role in S’s systematic subordination”. (ibid: 235)

While it is clear that this definition *as such*, has a normative, stipulative aspect, one might still ask: Has Haslanger really *normatively engineered* the concept of “woman” in relation to (today’s) ordinary usage?

Firstly, Haslanger admits that with her analysis she wants “to be responsive to some aspects of ordinary usage” and that the “goal of the project was not to capture *what we have in mind* when we use the terms ‘woman’, or ‘man’ [...] to describe ourselves or others” (Haslanger 2020: 235). But neither is this the aim of Wittgenstein’s descriptive conceptual investigations when we talk e.g. about sensations: If we were asked as ordinary people – and not as philosophers – what “sensations” are, it would probably not occur to anyone to answer that they are “grammatically private” (cf. PI 2009: 248). Wittgenstein’s descriptions of the rules of usage in ordinary language derive their purpose and meaning from the philosophical problems associated with these terms (cf. PI 2009: 109), i.e. they do not merely *copy* language usage but rather emphasise or *re-describe* the philosophically important aspects of our ordinary usages that we overlook in philosophising because of their ordinariness and simplicity.

In *this* sense of “description”, one can also recognise a “stipulative” moment in Wittgenstein’s descriptions, if ‘stipulative’ here means: “*This* is the

phenomenon we need to be thinking about” (Haslanger 2000/2012: 224); likewise, Haslanger’s approach to gender can also be understood as descriptive in this sense: Even if her account has a normative *claim*, Haslanger herself says in retrospect that with her analysis of “gender” she “revealed features of our meaning that we were mostly unaware of” and that these “might provide a better account of what we actually mean in dominant contexts than what we take ourselves to mean” (Haslanger 2020: 236 f.). Reformulated in more descriptive terms, her definition reads as follows: In ordinary language, we use the concepts “woman” and “man” not only as disinterested observations of different anatomies, but we use them – knowingly or not! – as designations for the complex relationships between certain physical characteristics and social-hierarchical positions. From a metaphilosophical perspective, this descriptive reading of Haslanger’s approach doesn’t seem so far removed from Wittgenstein’s previously mentioned example of the society, whose rules and hierarchies are highlighted by philosophy through a new seating arrangement (cf. BT 2005: 415). Haslanger’s analysis is thus less normative than she had originally intended: it does not *interfere* with the actual, operative concept used in ordinary language. Hence, her analysis rather fulfils the negative-critical function of (re-)descriptive approaches (cf. *ibid.*: 237): For, based on this description of how “woman” is used, one can criticise from a feminist perspective that, according to these rules of use, certain people would not be described e.g. as “woman” although they *should be*.

Without seeking to generalise, I think this shows that the boundary between descriptive and normative approaches is neither as clear nor as fundamental as is sometimes assumed in conceptual engineering (cf. Cappelen 2020: 133 f.), and thus other approaches of feminist conceptual engineering can also be understood as descriptive, at least in the sense, in which Wittgenstein’s philosophy is descriptive (e.g. Manne’s analysis of the concept of misogyny; cf. 2017: 63).

But what about approaches that really ameliorate, change and thus interfere with our everyday concepts? In “What is Sexual Orientation?” (2016), Dembroff suggested revising the currently common concepts of sexual orientation (heterosexual/queer), as these are not only inconsistent and imprecise but also assume binary sex or gender categories excluding non-conforming gender (cf. *ibid.*: 2). Their proposal is – very briefly – a socio-

politically improved version of our everyday concept of sexual orientation, according to which one's own sexuality is independent of one's own gender (cf. *ibid.*: 18 f.). According to this framework, e.g., “a cisgender man and a transgender woman disposed to sexually engage only with cisgender women have the same sexual orientation” (*ibid.*: 19).

However, Dembroff's new conceptualisation faces the challenge of implementation of these new concepts, i.e. the problem that the meaning of a word cannot simply be changed by philosophical arguments – no matter how good they are! – but is determined by numerous external factors (cf. Cappelen & Plunkett 2020: 3). While it is not impossible for philosophical investigations to change everyday concepts, this seems comparatively rare. Moreover, the conceptual change itself, i.e. the successful implementation of the altered term, doesn't depend on the philosophers themselves: An individual *cannot* control the meaning of terms. This is not just a metaphilosophical statement about philosophy's limits, but rather one of the substantive points Wittgenstein demonstrated in PI. For, if our concepts and their meaning are constituted by the rules of usage in everyday language-games, this means that these rules function *as rules*, i.e. normatively, only if they are embedded in a practice of human action, i.e. in customs, regularities and institutions (cf. PI 2009: 199, 206, 241). It is ultimately the agreement of people *in deed*, in their form of life, that determines which terms we use and how (cf. PI 2009: 241), not an individual philosopher.

3.

In section 2, I argued that while some approaches to conceptual engineering, in what they *actually* do, are at least not as far removed from Wittgenstein's philosophizing as it initially seemed, there are substantive arguments concerning the functioning of our concepts and language, which, coming from Wittgenstein, represent a serious challenge to ameliorative projects, such as Dembroff's (2016).

However, the main differences between Wittgenstein's philosophy and approaches in feminist conceptual engineering, I think, is in the *kinds* of concepts each analyses: Although both approaches are concerned with ordinary concepts, Wittgenstein was primarily interested in *philosophical*

concepts, in a narrower sense. For, concepts like “knowledge”, “thinking”, “I”, “mind”, “meaning” (etc.) can pose philosophical problems, when we interpret them in a way detached from their use in ordinary language: Without practical application, however, these terms no longer take hold; they are “idling wheels” and no longer belong to the machinery of “living language” (cf. Ms 137: 107a): “The confusions which occupy us” – i.e. Wittgenstein – “arise when language is, as it were, idling, not when it is doing work” (PI 2009: 132). Opposed to this, the feminist philosophers in conceptual engineering considered here are interested in social concepts and the moral or socio-political problems they pose. These problems do not arise when concepts are idling and language “goes on holidays” but rather – to stay with Wittgenstein’s image – when there is too much friction, when the concepts run hot and language burns out. I.e., these problems arise when certain social concepts do no longer fit with our life patterns. These are conceptual problems of a completely different, more practical nature than those that preoccupied Wittgenstein, and consequently require a different “therapy”, which may involve interfering with ordinary language.

However, the treatment of socio-politically problematic concepts was simply not Wittgenstein’s ‘cup of tea’: He himself said that a reform of language “for particular practical purposes, an improvement in our terminology to prevent misunderstandings in practice, may well be possible. *But these are not the cases we are dealing with*” (PI 2009: 132; *emph. JT*). The cases Wittgenstein is dealing with are the problems that arise when ordinary language is applied metaphysically in philosophical reflections, so the philosophical treatment of these problems is to bring back the words to their everyday use, where they are at home. *This* is the reason why philosophy in his sense must not interfere with the actual use of language but this does *not* mean that Wittgenstein therefore assumed that the words and concepts of our ordinary language are the only correct or reasonable ones or the ones most appropriate to the phenomena or do correspond to “our nature” (cf. BT 2005: 233; RPP II 1998: 426 f.; PPF 2009: 366; LW II 1999: 46): “For a world of a different kind” – and this may only be a different lifeworld within our world – “one would find the use of different linguistic instruments natural” (MS 137: 61b).

But what does that mean for actual conceptual change and philosophy’s role in it?

Firstly, it is noteworthy that Wittgenstein assumed that language and concepts change when people's lives or patterns of life change:

A pattern of life serves as the basis of a word usage. The pattern changes. The language-game begins to falter. (MS 167: 16r).

As can be seen from this, our concepts change “bottom up”, so to speak, when life changes, because life patterns, i.e. the interweaving of language-games, customs and situations, are the basis of our concepts.

Hence, what philosophy as conceptual work can do for conceptual change is indeed limited. Rather the impetus for and the realization of changes in our social concepts, i.e. the *actual* conceptual engineering, seems to lie with activists, social and political movements and institutions. However, I nevertheless think that philosophy can play an important role in such bottom-up conceptual change, because the particular difficulty of *conceptual* injustices is the internal connection between concepts and *thinking* (cf. RPP II 1998: §678 f.; Trächtler 2023: 7). Philosophy can help to make clear those things that we don't see because they are always before our eyes, i.e. it can help to make the prejudices in our concepts – be they grammatical or political – visible and thus wake us from the “hypnosis of familiarity” (cf. MS 137: 9a).

If you always represent everything by contours, you cannot represent a gradual transition. If one were used to representing all whole objects in straight lines, it would be difficult to free oneself from the prejudice that a curvilinear one could be something other than a fragment. (MS 134: 2 f.)

The “hypnosis of a familiar way of representation” – in Wittgenstein's image: in contours or straight lines – makes it difficult to free oneself from the prejudices that come with this mode of representation and to see the gradual transitions and the curving lines as a whole. In this sense, an ameliorative project such as Dembroff's can be helpful in clarifying certain problems and prejudices in our concepts of sexual orientation. However, as an ameliorative project that strives for top-down change of everyday concepts, it will fail at the hurdle of implementation. From a Wittgensteinian point of view the *emergence* of the implementation problem seems indeed symptomatic of “top-down”

conceptual engineering and thus of Ideal Language Philosophy's dogmatism that some concept *must* correspond to reality (cf. PI 2009: 130 f.). *This* would be the "typical philosophical" error that creates idling wheels:

The engineer's error is related to a philosophical error. The construction looks exactly like a machine and yet is not one. We have robbed it of any possibility of movement ourselves and do not know it. (MS 114: 118r)

The "usefulness of philosophy" consists precisely in seeing such prejudices or injustices in our concepts and understanding how they function in language use (cf. MS 133: 46e). This alone may not lead to an actual change of our concepts, but it is or can be an important component of feminist and other social movements and activism. Because if you want to move in order to achieve a certain goal, you have to know your way around. Philosophy can provide the necessary orientation with a map of the conceptual landscape, because such a map is not just a mirror-image of the landscape but can also contain an evaluation of the viewpoints, previously unexplored paths and bad routes. But maps alone cannot change the landscape. To use a Kantian echo, activism without clear concepts would be blind, and a feminist philosophy without activism would be empty.

References

- Cappelen, Herman (2018) *Fixing Language - An Essay in Conceptual Engineering*, Oxford: OUP.
- Chalmers, David (2020) "What is Conceptual Engineering and What Should it Be?", *Inquiry*, 1-18.
- Dembroff, Robin (2016) "What Is Sexual Orientation?", *Philosophers' Imprint* 16:3, 1-27.
- Haslanger, Sally (2000/2012) "Gender and Race – (What) Are They? (What) Do We Want Them To Be?", in: Sally Haslanger: *Resisting Reality*, Oxford: OUP, 341-364.
- Haslanger, Sally (2020) "Going On, Not in the Same Way", in: Alexis Burgess, Herman Cappelen and David Plunkett (eds.) *Conceptual Engineering and Conceptual Ethics*, OUP, 230-260.
- Horwich, Paul (2012) *Wittgenstein's Metaphilosophy*, Oxford: Clarendon Press.
- Manne, Kate (2017) *Down Girl: The Logic of Misogyny*, Oxford: OUP.
- Trächtler, Jasmin (2023) "Facts, Concepts and Patterns of Life – Or How to Change Things with Words", *Philosophies* 8:4, 58, 1-13.

Wittgenstein, Ludwig (1986) *Philosophical Investigations*, P.M.S. Hacker and Joachim Schulte (eds.), G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte (transl.), Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1998) *Remarks on the Philosophy of Psychology Vol. I & II*, G. H. von Wright and Heikki Nyman (eds.), C.G. Luckhardt and M.A.E. Aue (transl.), Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1999) *Last Writings on the Philosophy of Psychology Vol. II*, G. H. von Wright and Heikki Nyman (eds.), C.G. Luckhardt (transl.), Oxford: Basil Blackwell.

Wittgenstein, Ludwig (2005) *The Big Typescript*. C. Grant Luckhardt and Maximilian Aue (eds. And transl.). Oxford: Basil Blackwell.

Wittgenstein, Ludwig (2016) *Interactive Dynamic Presentation (IDP) of Ludwig Wittgenstein's Philosophical Nachlass*, Alois Pichler (ed.), <http://wab.uib.no/transform/wab.php?modus=opsjoner> (accessed: 12.12.2023), my translations, Bergen: WAB.

Three (Argumentative) Cheers for Historiographic Realism

Aviezer Tucker (Ostrava, Czechia)

Abstract

I present three arguments for a realist interpretation of historiography and the historical sciences, arguing that the historical sciences infer information about the past that is probably true and is constrained by and reflects what really happened in the past (Timmins 2022). One argument relies on the social epistemology that distinguishes historiography from art, and ideology. The second, demonstrates the differences between historical counterfactuals and “factual” historiography and reduces those differences to their full vs. truncated evidential foundations. The third argument relies on the history of historiography, on an “optimistic induction,” from the history of the historical sciences, which, unlike physics, have gone through a single founding scientific revolution, and have since expanded without revolutions that overturn previous paradigms.

I present three arguments for a realist interpretation of historiography and the historical sciences, arguing that the historical sciences infer information about the past that is probably true and is constrained by and reflects what really happened in the past. (Timmins 2022) One argument relies on the social epistemology that distinguishes historiography from art, and ideology. The second, demonstrates the differences between historical counterfactuals and “factual” historiography and reduces those differences to their full vs. truncated evidential foundations. The third argument relies on the history of historiography, on an “optimistic induction,” from the history of the historical sciences, which, unlike physics, have gone through a single founding scientific revolution, and have since expanded without revolutions that overturn previous paradigms.

1.

If historiography is art, fiction, or ideology, any truth to it is coincidental. Hayden White (1987, 1992) defended the view that much of historiography is ideological fiction, and different stories about the past should be judged according to aesthetic or ideological criteria: “[H]istorical narratives... are verbal fictions, the contents of which are as much invented as found and the forms of which have more in connection with their counterparts in literature than they have with those in the sciences.”(White 1978, 82).

This anti-realist concept of historiography is at odds with the social structure of the discipline. In art and ideology there are social “bubbles” of likeminded people who share aesthetic or political values. Members of these groups, say impressionists and abstract expressionists, theocrats and liberals, disagree but do not reason to convince each other; if they try they usually fail—people change artistic or political identities as a result of “conversions,” not convincing. By contrast, there are no historiographic social “bubbles.” most historians agree on most of their epistemic historiographic outputs, and even more so on their epistemic inputs, the evidence. This creates a space for debates about reasoning from evidential inputs to historiographic outputs. The broad historiographic consensus among very different people has emerged when historiography underwent a scientific revolution that many associate with Ranke and his methods, but preceded him in the late 18th century. This degree of consensus resembles the history and social structure of science more than those of art or ideology. (Tucker 2004)

The scientific community, in comparison with non-scientific communities, is marked by consensus that follows expertise. For example, on average scientists agree on Darwinian Evolution and Climate change more than any group of non-scientists who share the socio-economic and historical contexts of the scientists. The explanation of this gap is that scientists possess expertise, special knowledge that lay people do not always share. (Tucker 2014) Similarly, historians on average agree on historiography more than lay people who share their socio-economic and historical contexts.

Broad consensus that distinguishes a community of experts from a socially comparable “control group” can have three and only three explanations: shared biases of the experts, random agreement, and shared specialized knowledge. For example, in an opinion survey that asks: “Do you believe that massive public investment in historiographic research and education will greatly improve civic virtue and discourse?” Undoubtedly, historians would reach close to a consensus, agreeing with this statement. This consensus would far exceed the average agreement on this statement in a control group of non-historians. But the best explanation of this consensus would be professional bias—any profession would tend to agree that massive transfer payment to its members is beneficial to society. Those who would have to foot

the bill for this largesse may be less enthusiastic... Agreement can also be random. If a groups of historians go to a restaurant and spontaneously order the same item from the menu, this agreement may reflect neither bias nor expertise. One historian may like the dish, another may consider it the healthiest, yet another may have read a recommendation on a website, and yet another was instructed by his doctor to order that meal. This consensus of historians has neither a reason nor a bias; it is random. Finally, historians may agree because they share access to the same public evidence and they agree on the reasoning, the epistemic processes, that transmit and generate knowledge from the evidence, whereas non-experts do not know the evidence or understand the reasoning processes that the experts utilize.

If historians reach consensus because of biases, they must either have the same biases, or different biases should somehow converge to generate the same historiography. Historians have very different non-professional interests, passions, identities, institutional affiliations, and so on. So, they have no common biases that can correlate with the historiography they agree on more than control groups of non-historians would. The identities and interests of historians who agree on much of historiography are in too much conflict to plausibly generate historiographic consensus as unintended spontaneous result of different biases. The historiographic consensus is also too consistent and broad to plausibly follow random agreement. By elimination then, expert knowledge based of common evidence and reasoning explains the broad historiographic consensus and the gap between expert historiographic knowledge and lay opinions.

Still, Hayden White, among others, would retort that rationality, reason, is itself a type of ideological bias: “For subordinate, emergent, or resisting social groups, this recommendation—that they view history with the kind of ‘objectivity,’ ‘modesty,’ ‘realism,’ and ‘social responsibility,’ that has characterized historical studies since their establishment as a professional discipline—can only appear as another aspect of the ideology that they are indentured to oppose.” (White 1987, 81) White (and others) reduced “historiographies” to two opposing ideological biases. Reason and professional norms of historiographic research are associated with “centers of established political power and social authority” and opposed to the alternative bias of “subordinate, emergent, or resisting social groups.” This Manichean world

view shares the groundless oligarchic myth perpetuated by Plato and Aristotle that considered the lower classes and slaves irrational and hence unable to govern themselves, while holding the elite to be the paragon of rationality. It ignores, for example, the Enlightenment tradition that used reason to undermine irrational hierarchies and social order.

The reduction of historiography to an ideology can acknowledge that the historiographic community is heterogeneous in many respects, yet, arguably it is homogenous at least in sharing *cognitive values*. Expertise is considered then just an intermediate variable between shared cognitive values and correlated beliefs. Cognitive values determine which statements are worthy of being considered knowledge. If shared cognitive values are necessary for forming beliefs, the historiographic expertise hypothesis may have to be qualified as relative to a particular bias, a shared set of cognitive values.

Consensus on cognitive values, just like the consensus on the historiographic outputs that result from it, can result from bias, coincidence, or expertise. The emergence of historiography as a science was accompanied by a shift from cognitive values that considered knowledge of the past to follow faith, revelation, ancient wisdom and above all tradition, to the current evidentiary and critical reasoning values. These new values were adopted because historians believe they were more conducive to the discovery of more truths about history. The “conduciveness to knowledge and truth” of cognitive values hypothesis can be tested just like the historiographic expertise hypothesis, by comparing it with bias and coincidence hypotheses. Historians, as experts with special knowledge, may correlate more strongly with rational cognitive values because they are conducive to the attainment of knowledge than comparable control groups that are not in the business of obtaining knowledge of the past have fewer reasons to adopt. The competing hypotheses that claim to explain a significant correlation of historical experts with cognitive values by any other (particularly external social or cultural) biasing variables would find it quite difficult to explain the appeal of these values to very different experts and their lower appeal to very similar control groups of non-experts. The high correlation between historians (and other scientists) and shared cognitive values in comparison with a comparable “control” group of non-historians (or non-scientists) may be explained then by the conduciveness of these cognitive values to knowledge (Tucker 2014), in comparison with

alternative values, such as those that value historiographies according to their therapeutic effects, whether they help their audiences feel better, less guilty (e.g. Holocaust denial), more proud and self-confident (nationalist and other identitarian historiographies), or belonging to an old and venerable tradition (anachronistic-identitarian) that go back to what Vico analyzed as “conceit of the nations,” the claim for primacy in the invention of civilization founded on national memory that stretches back to the dawn of civilization. (Vico 1984, paragraph 125). Rational historiography, based on evidence and reasoning, can also have therapeutic effects, but the cognitive values that determine historiography are primary and indifferent to their therapeutic effects.

To use Windelband’s (2015) distinction between *origins*, the historical context of emergence, development, and reception, of ideas, and *validities*, what make the outcomes of origins acceptable or not, historiographic reasoning from the evidence validates historiography irrespective of its origins. The existence of origins does not have to bias the validation process. For example, Marxism is among the main origins of the sub-discipline of Economic History because it drew attention to the historical significance of economic structures and change even prior to the Industrial Revolution. However, the validity of economic historiography has nothing to do with this origin.

Some philosophies of historiography double down on the denial of historiographic realism by combining “empiricism” with aestheticism. Ankersmit (1995, 2001, 2012) interpreted historiography as narratives constructed on empirical facts. Ankersmit divided historiography between descriptive “empirical” propositions about individual historical facts, and narratives that are about themselves as much as about history. Artistic representations may describe an object like the Houses of Parliament in London or the Cathedral in Rouen, but may also represent it with a particular style (impressionist, expressionist, and so on) that is about the art work itself rather than the world. As much as aestheticians distinguish different painters or artistic styles that represent the same object, historiographic connoisseurs may distinguish equally adequate historiographic narratives. In Ankersmit’s philosophy, the “empirical” evidence confers on historiographic narratives only *adequacy*. Evidence may be shared by contradictory but equally adequate narratives. The historiographic narrative contains true ingredients, but on the whole is neither true nor false because it is art. Composing historiography is

analogous then to baking a strudel: Obviously, the process begins with apple pickers choosing which apples are ripe and which are rotten. The apple pickers are necessary and their work is important; the greatest chef cannot bake from rotten apples. But once the ripe apples are brought to the kitchen, the cooks can display their genius: Cooking apprentices study the kitchen and not the orchard. Likewise, historians pick and choose the “ripe” facts and discard the forged or unreliable ones. They bring the collected “facts” to the historiographic “kitchen” where historians conceptualize, slice and combine them with added explanations, value judgements, a pinch of theory for flavor, pour them into readymade narrative pans to give them structure and form, and slide them into the narrative oven to bake together, until ready for public consumption. Different cooks may make different dishes and give them different tastes in different shapes from the same types of apples and none of the dishes is necessarily better than the other. It is ridiculous to reason for or against dishes unless the apples are rotten: *De gustibus non est disputandum*.

Still, making historiography is not analogous to the baking of strudels. The past is unobservable in principle. The historical train has always left the station before the historian could arrive. Historiography cannot be an empirical science that observes its objects of study. At most it can observe the evidence, and it is mostly testimonial, the subject of the epistemology of testimony. There are no ripe and ready facts in the archives that can be picked by observation without reasoning. The archive is not made of apple-like distinct atomic observational facts, ready to be baked in the historian’s narrative workshop. Historiographic “facts” are not the evidence; historiographic facts just like the rest of historiography has to be inferred from the evidence in a process of historiographic reasoning. The final product of historiographic research, especially one directed for popular consumption, indeed may have the form of a narrative. But that final form ready for consumption is the *superstructure* of historiography as Goldstein (1976) put it, the final sliced and “garnished with cream” on a silver plate form of the strudel, a dish whose substance is reasoning from the evidence.

2.

If historiographic realism is right, historiographic counterfactuals must have an ontological status different from that of historiography. If the two are indistinguishable, historiographic realism fails.

Distinguishing the two is not obvious. Though some historians claim that historiographic counterfactuals are just literary fiction or “parlor games,” some historiographic counterfactuals can have similar degrees of justification and certainly to those of factual historiography, for example, counterfactuals about historical succession: Had Donald Trump been relieved of his presidency, Mike Pence would have become the 46th president of the United States. Had Lee Harvey Oswald missed J. F. Kennedy, L. B. Johnson would not have succeeded him then as the President of the United States. These are highly probable determined counterfactuals. It may be argued, following David Lewis’ (1973) counterfactual theory of causation, that counterfactuals are indistinguishable from historiography because whenever historians make causal judgements, they assume, at least implicitly a historiographic counterfactual: Had the cause not occurred, nor would have the effect. The evaluation of counterfactuals, according to David Lewis’ theory, is derived from historiographic factual reasoning because historiographic counterfactuals are justified by measuring their resemblance to factual historiography: The counterfactual most similar to factual historiography is the most plausible.

A useful starting point may be a comparison of their evidential premises with the evidential premises of factual historiographic reasoning. The inputs to historiographic counterfactual must both subtract from, and add to, the evidence that factual historiography uses as inputs. Historiographic counterfactuals must presume by *fiat* imaginary evidence that determines the antecedent of the counterfactual. We can call this evidence “ghostly” because it is not specified but must be presumed to infer that J. F. Kennedy was not assassinated, or that Trump did not complete his term of office. The ghostly evidence must be consistent with the rest of the evidential inputs for historiographic counterfactuals. Factual historiographic evidence necessarily, by definition, contradicts and disproves all historiographic counterfactuals: There is plenty of reliable evidence that J. F. Kennedy was assassinated and succeeded by L. B. Johnson, and Trump not only completed his term of office,

but even sought to extend it further. Still, for the historiographic evidence to serve as input for a counterfactual output rather than to refute it and to fit the “ghostly” evidence, the evidence against the counterfactual must be *truncated*, we agree to suspend belief in part of the evidence. Counterfactual reasoning suspends belief in historiographic evidential inputs that contradict the outputs of the counterfactual. The counterfactual then is a function of the truncated but consistent evidence.

Historiographic counterfactual reasoning depends on whether there is sufficient evidence, following the truncation of evidence that contradicts ghostly counterfactual evidence, to infer the consequents of the counterfactual. This depends both on how much historiographic evidence was there to begin with, and on how much of it must be truncated to avoid incoherence with the “ghostly” counterfactual evidence. When much of the evidence is truncated, there is not much left as inputs to infer the consequent of the counterfactual from the evidence. Consequently, the counterfactual becomes fiction.

3.

Scientific anti-realism that takes science to be merely the reflection of its historical origins rather than independent reality assumes historiographic realism to establish its historicist thesis, partly because a realist interpretation of historiography is more plausible than a realist interpretation of physics.

The anti-realist “pessimistic induction,” infers from historical scientific revolution that demonstrated that scientific theories were false or inadequate that *all* scientific theories *will* be proven false or inadequate. (Laudan 1981, Wray 2015) However, contemporary historiography and the historical sciences, by contrast, can support an “optimistic induction” that infers that it is probably true and its main conclusions will probably not be overturned. Unlike physics, historiography and historical sciences like comparative linguistics, geology, and evolutionary biology had one and only one, founding scientific revolution. Since the revolutions associated with Ranke in historiography, Lyle in geological history, and Darwin in biological history, the historical sciences have made progress, but have not overhauled the founders’ representations of the past. We know more about the human past than Ranke

did and much more about the biological past than Darwin did, but the modern historical sciences have not overturned their theories and conclusions. Though there are many new methodologies, historians still go to the archives to compare primary sources and biologists still study fossils and assume natural selection. Since the founding of these sciences by Ranke, Darwin, Lyell, Bopp, and so on, they have expanded progressively but not revolutionarily by adding new theories and methods that led to the discovery of new evidence, and new reliable inferences about the past. Historiography and the historical sciences are internally coherent and have not accumulated anomalies that may prepare expectations for a scientific revolution. A second scientific revolution in historiography, comparable to those physics has undergone, would require historiography to replace its ontology with ontologies as different as those of Hegelian idealism or of conspiracy theories, not impossible, but unlikely.

References

Ankersmit, Frank (1995) "Statements, Texts and Pictures", in: Ankersmit, Frank and Kellner, Hans eds. *A New Philosophy of History*, London: Reaktion Books, 212-240.

Ankersmit, Frank (2001) *Historical Representation*. Palo Alto CA: Stanford University Press.

Ankersmit, Frank (2012) *Meaning, Truth, and Reference in Historical Representation*. Ithaca NY: Cornell University Press.

Goldstein Leon J., (1976) *Historical Knowing*. Austin: University of Texas Press.

Laudan, Larry (1981) "A Confutation of Convergent Realism", *Philosophy of Science*, 48, 19–48

Lewis, David (1973) *Counterfactuals*, Oxford: Blackwell.

Timmins, Adam (2022) *Towards a Realist Philosophy of History*, Lanham, Md.: Lexington Books, 2022.

Tucker, Aviezer (2004) *Our Knowledge of the Past: A Philosophy of Historiography*, Cambridge: Cambridge University Press.

Tucker, Aviezer (2014) "Epistemology as a Social Science: Applying the Neuman-Rubin Method to Explain Expert Beliefs", in: Carlo Martini & Marcel Boumans eds., *Experts and Consensus in Social Science*, Dordrecht: Springer, 2014, 155-170.

Vico, Giambattista (1984) [1744] *The New Science*, trans. Thomas Goddard Bergin & Max Harold Fisch, London: Cornell University Press.

White, Hayden (1978) *Tropics of Discourse*, Baltimore: Johns Hopkins University Press.

White, Hayden (1987) *The Content of the Form*, Baltimore: Johns Hopkins University Press.

White, Hayden (1992) "Historical Emplotment and the problem of Truth", in: Saul Friedlander ed., *Probing the Limits of Representation: Nazism and the 'Final Solution'*, Cambridge MA: Harvard University Press, 37-53.

Windelband, Wilhelm (2015) "Critical or Genetic method (1883)", in: Sebastian Luft (ed.), *The Neo-Kantian Reader*, New York: Routledge, 271-286.

Wray, B. K. (2015) "Pessimistic Inductions: Four Varieties", *International Studies in the Philosophy of Science*, 29, 1, 61-73.

Not All Kids Get Medals

Bram Vaassen (Umeå, Sweden)

Abstract

This paper develops a new argument for moderate causal competition between events and their grounds. The starting point is Lee's recent observation that shifting the focus from the grounding relations between causal relata to the grounding relations between causal relations themselves reveals that higher-level causal relations sometimes ground lower-level causal relations, rather than vice versa (Lee 2021). Lee argues that such grounding relations eliminate the need for causal competition entirely. I argue that they highlight the need for a moderated brand of such competition instead. In particular, I argue that downward grounding relations between causal relations will entail both implausible causal claims and implausible grounding claims unless there is moderate causal competition between grounded event and their grounds.

The new argument for causal competition innovates work on causal competition on two fronts. First, it shifts the focus from the metaphysical relations between causal relata to the metaphysical relations between the causal relations themselves. Second, the moderate brand of causal competition that it recommends contradicts standard difference-making accounts (e.g., Lewis 1973, Woodward 2021), exclusion principles (e.g., Kim 2005, Baumgartner 2018), and proportionality principles (Yablo 1992a,b, List and Menzies 2009, Zhong 2020a,b).

1. Introduction

Most events ground, and are grounded by, other events. For example, my backpack being red grounds its being coloured and is grounded by its being crimson. Similarly, the Gettysburg Address occurring grounds *some* address occurring, and was grounded by a collection of microphysical events occurring at the same time and place. Studying such grounding relations is taken to be a central task in metaphysics (cf., Schaffer, 2009), and studying how they affect the causal status of the events involved has taken a central place in the philosophy of causation (e.g., Kroedel and Schulz, 2016).

Lee (2021) carves out a new and intriguing position in the debate on causal relations and grounded events. In a nutshell, Lee argues that, even when an event *a* is grounded by an event *b*, the causal relation between *b* and an effect *e* might still be grounded in the causal relation between *a* and *e*. As Lee puts it: 'Causation involving more fundamental events is not always itself more fundamental — indeed, sometimes it is less fundamental' (Lee, 2021, p.

2). For example, even if my thirst is grounded in neural event n , it might still be the case that n causing me to drink is grounded in my thirst causing me to drink. Call this view ‘Downwardism’.

Downwardism provides a principled answer to a question that has often gone unasked: how do causal relations ground one another? Given recent scholarly attention to grounding, causation, and their interrelation, this lacuna is surprising. However, when breaking new ground, it is important to get things right. Lee misjudges the impact of Downwardism. He concludes that Downwardism eliminates the need for causal competition between events that stand in grounding relations. Much to the contrary, Downwardism provides us with new reasons to endorse causal competition, albeit in a moderate form. Without a place for such moderate competition in our theories of causation, we are forced to endorse both implausible causal claims and implausible grounding claims.

I will make my case as follows. In §2–3, I spell out Downwardism and the no-competition view in more detail. In §4, I argue that this combination of views encounters serious problems and in §5 I propose a solution in terms of moderate causal competition.

2. Downwardism

Grounding is a metaphysical brand of asymmetric determination that backs explanation and tracks relative fundamentality. Phrases like ‘because’ and ‘in virtue of’ provide us an intuitive grip on such relations. For example, there is a cup of coffee on the table *because* of, or *in virtue* of, there being fundamental particles arranged cup-of-coffee-on-table-wise. The arrangement of the particles grounds the coffee cup being on the table. Consequently, the arrangement is more fundamental than the cup of coffee and the table, and it explains why there is a cup of coffee on the table. Following convention, we can say that grounded events are ‘higher-level’ and their grounds are ‘lower-level’.

It is natural to assume that the direction of grounding between causal relations aligns with the direction of grounding between the causal relata involved. For example, one would expect that higher-level causal relations, i.e.,

causal relations involving grounded events, are grounded in their respective lower-level causal relations, i.e., the causal relations involving their grounds, rather than vice versa. However, our use of ‘because’ and ‘in virtue of’ indicates that this is not the case.

To see this, consider the following example presented by Lee (2021, p. 9). Suppose Xanthippe hated her husband Socrates so much that she would be happy to be relieved of him. Suppose also that being a widow in Ancient Greece was such an undesirable social position that it would cause anyone, even Xanthippe, anguish. The following now seems true:

- (1) Socrates’ death caused Xanthippe to be in anguish because it made her a widow.

Here is a natural interpretation of the scenario. The causal relation between Socrates’ death has two main building blocks: (i) the grounding relation between his death and Xanthippe’s being a widow, and (ii) the causal relation between her being a widow and her anguish. But this natural interpretation also means that the lower-level causal relation between Socrates death and Xanthippe’s anguish is grounded in the higher-level causal relation between Xanthippe’s becoming a widow and her anguish (cf., Fig 1). We now have an apparent example of a higher-level causal relation grounding a lower-level causal relation.

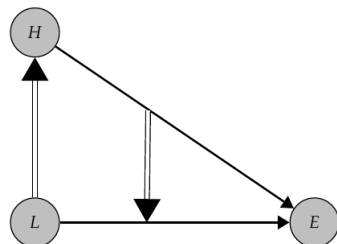


Figure 1: Nodes represent events, single arrows represent causal relations, and double arrows represent grounding relations. The higher level event (H) is grounded in the lower level event (L) and both cause the effect (E). L causing E is grounded in H causing E (cf., Fig 5. in Lee (2021)).

This pattern is easy to replicate. Martha Stewart selling certain stocks caused her to be convicted because it was a case of insider trading. Her selling the stocks grounds her committing the crime of insider trading, but the causal relation between the crime and the conviction grounds the causal relation

between the sales and the conviction (cf., Lee, 2021, p. 9–10). My driving 60 mph caused me to get a ticket because it is was a case of speeding. The driving 60 mph grounds the speeding, but the causal relation between the speeding and ticketing grounds the causal relation between the driving 60 mph and the ticketing. And so on. Upon closer scrutiny, the lower-level causal relations in these cases appear to be grounded in (i) the grounding relation between the lower-level event and the higher-level event and (ii) the causal relation between the higher-level event and the effect.

Based on this pattern, we can make the following claim (cf., Lee, 2021, p. 17):

Downwardism f caused g is grounded in p caused q if (i) f caused g and p caused q , and (ii) $g = q$ and f grounds p

Importantly, Downwardism posits only sufficient conditions for grounding causal relations and thus explicitly allows for causal relations to be built in other ways as well. Even so, Downwardism is a significant claim in its own right.

For one thing, Downwardism answers a question that has been lurking in the background of higher-level causation debates. In these debates, there has been a strong focus on what relations should hold between two events such that they can unproblematically cause the same effect. Standard approaches have not considered the further question of how the metaphysical relations between events affects the metaphysical relations between the causal relations they enter into. Perhaps it was assumed that both kinds of metaphysical relations would align: if event e_1 grounds event e_2 , then any causal relation that e_2 enters into must be grounded in the causal relations featuring e_1 . As plausible as that sounds, cases like Xanthippe's widowing put pressure on this position, and anyone who wants to defend it now faces the challenge of explaining such cases without allowing higher-level causation to ground lower-level causation.

Moreover, allowing causal relations at different levels to stand in grounding relations helps combatting ontological parsimony worries about higher-level causation. To see how this works, we need to take a look at so-called debates surrounding 'causal competition'.

3. Causal Competition

Proponents of causal competition maintain that events in grounding relations compete for causal status relative to a given effect, such that one can exclude the other from causing that effect (cp., Yablo, 1992a,b; Kim, 1998; List and Menzies, 2009; Zhong, 2020b). Such causal competition is typically imposed to stop us from overascribing causal status. According to Lee, adopting Downwardism eliminates the need for causal competition by reducing the ontological cost of causal relations. I disagree. Downwardism reduces the ontological cost of causal relations, but it does not eliminate the need for causal competition. Let us look at Lee's arguments first.

One central motivation for being conservative when ascribing causal status is to avoid an overcrowded ontology. For example, many philosophers maintain that, if my thirst is grounded in neural state *n*, these events cannot both cause my drinking water without positing some ontologically costly extra causal relation (e.g., Kim, 1998; Zhong, 2020b). By having one of the events outcompete the other for causal status, one ensures that no such costly extra relations are added to the ontology.

However, if we allow higher-level causation to ground lower-level causation, having causation at different levels becomes ontologically cheap. The lower-level causal relations can be grounded in the higher-level relations, and grounded entities or relations impose no further costs on our ontology than the entities or relations they are grounded in (cf., Schaffer, 2015). So, if *Downwardism* is true, there seems to be no strong reason why events that stand in grounding relations cannot share causal status unperturbed. Both Socrates' death and Xanthippe's widowhood could cause Xanthippe's anxiety, without these two causal relations overcrowding our ontology. After all, the causal relation involving Socrates' death is grounded in the causal relation involving Xanthippe's widowhood and thus imposes no extra ontological costs. Lee concludes that causal competition between events that stand in grounding relations is 'but a mirage' (Lee, 2021, p. 27).

Allowing causal relations to ground one another thus allows us to buy causal relations at a metaphysical discount. However, this observation only eliminates the need for causal competition insofar as it is motivated by a need

for ontological parsimony. There are further reasons for believing that events and their grounds compete for causal status, and these reasons don't disappear by adopting *Downwardism*. Quite to the contrary, endorsing *Downwardism* provides us with further reasons to believe in causal competition.

4. The problem

Here is the problem. Without causal competition, one is forced to endorse implausible causal claims. Combined with *Downwardism*, these implausible causal claims result in implausible grounding claims. As the case for *Downwardism* rests on our intuitive judgments about grounding claims, we cannot embrace these implausible grounding claims without undermining the case for *Downwardism* as well. So, *pace* Lee, the intuitive case for *Downwardism* should motivate us to embrace causal competition. Or so I argue in the remainder of this text.

We can start by looking at some problem cases for no-competition views. Suppose Garfield is only grumpy on Mondays. Its being a Monday today grounds its being a weekday and causes Garfield to be grumpy. Even so, the following seems false:

(2) Its being a weekday causes Garfield to be grumpy.

After all, Garfield is only grumpy on Mondays. If it had been any other weekday, he would have been in fine spirits. Similarly, if Sophie, a pigeon who pecks exclusively at red objects, pecks at a red object, it holds that its redness grounds the object being coloured and caused her to peck, but the following seems false (cf., Yablo, 1992a,b):

(3) The object being coloured causes Sophie to peck.

Events related by a grounding relations do not always share causal status. However, accounts of causation that do not allow for causal competition between grounding and grounded events typically entail that (2) and (3) are true. For example, if we hold a simple counterfactual account, according to which *c* causes *e* if the absence of *c* would have lead to the absence of *e*, we are forced to accept such claims. After all, if it had not been a weekday, Garfield had not been grumpy, and if the pebble had not been

coloured, Sophie would not have pecked. These counterfactuals follow regardless of whether we make the antecedents true by inserting miracles, as per Lewis (1973), or by a hypothetical intervention that holds all else fixed, as per Woodward (2021).

Problem cases multiply once we consider events that share grounds. Suppose Ananya likes the bitter taste of Relativ Vodka. The taste of Relativ Vodka is grounded in the particular microphysical features of Relativ Vodka, which also happen to ground the fact that Relativ Vodka is transparent. If we assume that events related by grounding relations share causal status, we can make the following steps:

- (4) The bitter taste of Relativ Vodka causes Ananya to like it.
- (5) The microphysical features of Relativ Vodka causes Ananya to like it.
- (6) The transparency of Relativ Vodka causes Ananya to like it.

On the assumption that Ananya has no particular preference for transparent drinks, (6) seems false. However, given that their grounds overlap, the absence of the transparency would make for the absence of the grounds for bitterness as well, meaning that there is no reason to believe she would like Relativ Vodka if its transparency was absent. Because the absence of a grounded event presupposes the absence of its grounds, no-competition views typically allow events in grounding relations to share causal status without competition or strife. Defenders of no-competition views tend to bite the bullet on cases like (2), (3) and (6) and maintain that these claims are not so much false as misleading or unexplanatory (Bontly 2005, Weslake 2013, Kroedel 2020, Woodward 2021).

But, as downwardists, we cannot accept (2), (3) and (6) without committing ourselves to implausible grounding claims as well. Consider again the case of Sophie. The pebble's being coloured is grounded in its being red and Sophie pecking = Sophie pecking, thus satisfying (ii) in Downwardism. If we follow the no-competition view, we must accept that both being red and being coloured cause the pecking, and (i) is satisfied as well. Consequently, we would be forced to accept that the causal relation between its being red and Sophie's pecking is grounded in the causal relation between its being coloured and Sophie's pecking. In other words, we would be forced to accept:

(7) The object's being red caused Sophie to peck because its being coloured caused Sophie to peck.

(8) The object's being red caused Sophie to peck because it made it the case that the object was coloured.

Both claims seem patently false. Taking similar steps, they would be forced to accept:

(9) Its being a Monday caused Garfield to be grumpy because its being a weekday caused Garfield to be grumpy.

(10) Its being a Monday caused Garfield to be grumpy because it made it the case that it is a weekday.

(11) The microphysical features of Relativ Vodka cause Ananya to like it because its transparency caused her to like it.

(12) The microphysical features of Relativ Vodka cause Ananya to like it because it makes the vodka transparent.

All of which seem false as well.

Biting the bullet on all these cases would put Downwardism in an awkward position. Its ability to capture the intuitive force of grounding claims was supposed to be one of its main motivations. There must be a better way.

5 A solution

The obvious solution is to allow for causal competition. Such competition, it appears, did not only serve to keep our ontology lean. It also serves to get our causal claims right. The remaining question is how we can adjudicate the causal competition such that it fits *Downwardism*.

This is not an easy task. If *Downwardism* is to be an interesting claim, causal competition had better not be a cut-throat affair. Several accounts of causal competition only allow for one causal level, thus excluding any levels that ground, or are grounded in, the causal level (e.g., Kim, 1998; List and Menzies,

2009; Zhong, 2020a). On such views, two events that stand in a grounding relation cannot cause the same event, and conditions (i) and (ii) in *Downwardism* cannot be satisfied together.

We are thus in need of a more moderated causal competition, where events that stand in grounding relations can, but do not have to, cause the same effect. Several accounts of moderate causal competition are making the rounds, but they all build on the same observation: a higher-level event *h* is causally outcompeted by its grounding event *l* if *h* would not be followed by the purported effect if *h* had been grounded in a normal variation of *l* (e.g., Blanchard, 2020; Woodward, 2020; Vaassen, 2022). For example, its being a weekday will be followed by Garfield's being grumpy if it is a weekday in virtue of its being a Monday, but not if it is a weekday in virtue of its being any other weekday. On Thursdays, it is a weekday, but no grumpiness ensues. *Mutatis mutandis*, the same holds for the pebble's being coloured and Sophie's pecking, Relativ Vodka's transparency and Anyana's liking it, and other failures of higher-level causation: the higher-level event would not be followed by the purported effect if it had been grounded in a different lower-level event. We can say that these higher-level correlations are 'ground-sensitive'.

By contrast, the higher-level correlation in cases of higher-level causation is typically insensitive to changes in the grounding event. For example, the object being red still counts as a cause of Sophie pecking, because she would have pecked even if the redness had been grounded in a different shade, like crimson or auburn. Similarly, Ananya likes bitter drinks regardless of how the bitterness is realized microphysically, and Garfield hates Mondays regardless of which specific date of the year makes it a Monday. Finally, it is plausible that Xanthippe's widowhood would still have resulted in her anguish if she had married someone else and had subsequently lost that spouse. After all, the anguish was due to the precarity of being a widow rather than the loss of Socrates specifically. By demanding that higher-level causes would result in their effects even if they had been grounded in different events, the Downwardist could get just the right amount of causal competition. Such moderate competition isn't a case of winner takes all, but not all the kids get medals either.

Certainly, worries about moderate causal competition abound as well. Opponents worry that there is no way of adjudicating causal competition that systematically gets the cases right, and conclude that competition views confuse features of explanation for features of causation. It is an open question whether a moderate causal competition based on ground (in)sensitivity or similar criteria can address these objections satisfactorily, but the Downwardist had better hope that they can. If *Downwardism* is to be saved from seemingly false claims like (7)–(12), it needs a way to avoid (2), (3), and (6) as well. As far as I can see, allowing for some friendly causal competition is the only solution currently on offer.

References

- Baumgartner, M. (2018) "The inherent empirical underdetermination of mental causation", *Australasian Journal of Philosophy* 96(2), 335–350.
- Blanchard, T. (2020) "Explanatory abstraction and the goldilocks problem: Interventionism gets things just right", *British Journal for the Philosophy of Science* 71(2), 633–663.
- Bontly, T. D. (2005) "Proportionality, causation, and exclusion", *Philosophia* 32(1), 331–348.
- Kim, J. (1998) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, MIT Press.
- Kroedel, T. (2020) *Mental Causation: A Counterfactual Theory*, Cambridge University Press.
- Kroedel, T. and M. Schulz (2016) "Grounding mental causation", *Synthese* 193(6), 1909–1923.
- Lee, S. (2021) "Building low level causation out of high level causation", *Synthese* 199(3-4), 9927–9955.
- Lewis, D. (1973) "Causation", *The Journal of Philosophy* 70(17), 556–567.
- List, C. and P. Menzies (2009) "Non-reductive physicalism and the limits of the exclusion principle", *The Journal of Philosophy* 106(9), 475–502.
- McDonnell, N. (2017) "Causal exclusion and the limits of proportionality", *Philosophical Studies* 174(6), 1459–1474.
- Schaffer, J. (2009) "On what grounds what" In D. Manley, D. J. Chalmers, and R. Wasserman (Eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 347–383. Oxford University Press.
- Schaffer, J. (2015) "What not to multiply without necessity", *Australasian Journal of Philosophy* 93(4), 644–664.

- Vaassen, B. (2022) "Halfway proportionality", *Philosophical Studies*, 1–21.
- Weslake, B. (2013) "Proportionality, contrast and explanation", *Australasian Journal of Philosophy* 91(4), 785–797.
- Woodward, J. (2020) "Causal complexity, conditional independence, and downward causation", *Philosophy of Science* 87(5), 857–867.
- Woodward, J. (2021) *Causation with a human face: Normative theory and descriptive psychology*, Oxford University Press.
- Yablo, S. (1992a) "Cause and essence", *Synthese* 93(3), 403–449.
- Yablo, S. (1992b) "Mental causation", *Philosophical Review* 101(2), 245–280.
- Zhong, L. (2020a) "Intervention, fixation, and supervenient causation", *Journal of Philosophy* 117(6), 293–314.
- Zhong, L. (2020b) "Taking emergentism seriously", *Australasian Journal of Philosophy* 98(1), 31–46.

Acknowledgement

I am grateful to Karen Bennett, Jonathan Schaffer, Audrey Powers, Adrian Liu, Jordan Bridges, and Carolina Sartorio for comments on earlier drafts of this work.

This research was funded by a Swedish Research Council International Postdoc Grant [grant number 2019-0638].

A Response to Sider on Bare Particulars

Alice van't Hoff (Princeton, USA)

Abstract

Some philosophers have worried that there couldn't be bare particulars since such things would have to be possibly propertyless and nothing can be without being in some way. Ted Sider has argued otherwise: the nature of bare particulars, he suggests, could be being bare. I argue that if this were so, then the bareness of such particulars would need to be grounded in some way. I survey three possible ways that this demand could be met: by grounding a particular's bareness in its having no (sparse) properties, by taking its bareness to be zero-grounded, or by understanding the bareness of a particular as an ungrounded (i.e. fundamental) fact. I suggest, though, that none of these approaches succeed and therefore conclude that there is no adequate way to ground the bareness of bare particulars. It follows that, *pace* Sider, being bare isn't a way of being. The initial complaint against bare particulars is thereby upheld.

According to substratum theorists, the world is made up of universals and the particulars that realize them. These realizing particulars are sometimes called "bare particulars". (I'll precisify my use of this phrase below.) Some philosophers have thought that there was something objectionable about bare particulars (see citations in Sider 2006: n. 1 394). Ted Sider argues in his Sider 2006 that such concerns go awry. Against Sider, however, I hope to show that there really are reasons to be suspicious of bare particulars: very roughly, my argument is that nothing could ground the bareness of bare particulars.

1. Introducing the objection

Let's start by clarifying our terms: distinguish between *thin* particulars and *bare* particulars. According to substratum theorists, thin particulars are the instantiators of universals. They therefore have many rich and varied properties. *Bare* particulars by contrast are the *unempropertied* parts of thin particulars. (Thank you to Eliya Cohen for suggesting this term and for helping to clarify this distinction. Note that bare particulars should not be confused with Williamson's bare *possibilia*, see Williamson 1998) The proponent of substratum theory must accept that there could be thin particulars. It isn't obvious that they must also think that some particulars could possibly be *bare*—perhaps bare particulars are necessarily inseparable from the thin particulars that they compose.

Nonetheless, I will now suggest that substratum theorists *are* in fact committed to this possibility. In making this claim, I am relying on a Humean intuition according to which there can be no connections between distinct existents. If so, it follows that bare particulars should be separable from the thin particulars of which they are parts. Against this, Sider suggests that the foregoing line of reasoning would also generate uncomfortable problems for the substratum theorists' rivals (Sider 2006: 391–2). I think though that nominalists have a response to this kind of *tu quoque*. The worry is that the kind of combinatorial reasoning that would require us to conceive of bare particulars apart from the wholes of which they are part would also require that the nominalist acknowledge the possibility that "x is F" is false for any F. But nominalists have a well-motivated and non ad-hoc answer to this: the Humean principle to which I appealed is, they could suggest, restricted to the things that exist. Nominalists do not believe that any Fs exist. So there is no pressure to accept a combinatorial principle that would force the nominalist to accept the possible falsity of "x is F" for any F.

Let's assume then that substratum theorists must acknowledge the possibility that there could be something that is bare in the relevant sense of that term. A problem looms. For the opponent of substratum theory can argue that there cannot be things that are *actually* bare. Everything that there is must be some way, it must have a nature. But by definition bare particulars lack any of the properties that could constitute such a nature. So they have no such nature. Sider suggests that this argument rests on a "subtle confusion" (ibid.: 392–3). An entity, he claims, could have a nature by failing to have properties. Its nature would be its bareness, so to speak. There is something puzzling about this: it has it that a thing could have a way of being by being no way. But this is precisely what the opponent of bare particulars was suspicious about. Let's introduce the predicate of "being a bare particular" (we use superscripts here to indicate the "order" of a variable, i.e. 1st order variables range over individuals, 2nd order variables over predicates of individuals etc.): $\lambda x^2. x^2(y^1) \leftrightarrow \neg \exists z^2(z^2(y^1))$. This predicate cannot correspond to a feature of reality: for it would be instantiated if and only if it is not instantiated, and this is paradoxical.

According to Sider this way of thinking relies on a failure to distinguish between sparse and abundant conceptions of properties: according to the abundant conception, there is a property corresponding to every sensical predicate; according to the sparse conception, there are only properties corresponding to the fundamental joints of nature. (See discussion in Lewis 1986: 59—61.) Given this, the variable z in the foregoing predicate should be understood to range only over predicates picking out sparse properties, but x can range over predicates picking out abundant properties.

2. Responding to Sider

There is, however, a fly in the ointment: I will now argue that the Siderian proposal just outlined does not succeed. The basic idea is that if some bare particular p has an abundant property G , then there is a fact f that must be grounded (namely, the fact that p has G). The sparse conception is a minimalist picture in the sense that it implies that there are true predicative claims that do not pick out a (sparse) property. Since these claims are true, the minimalist should claim that their truth is grounded in some complex of facts about sparse properties. So f should similarly be grounded in this way. Against this, we might want to argue that f is somehow grounded by some facts about *abundant* properties. Yet if we are taking seriously the sparse/abundant distinction, then we cannot simply help ourselves to abundant properties in this way. It is true that Bilbo Baggins doesn't exist. But that fact isn't grounded in his realizing the property of non-existence. If Bilbo does have the (abundant) property of non-existence, he has it in virtue of his having *other* properties. But in that case, there is a problem for Sider's suggestion, since it seems that no sparse property could ground the attribution of the abundant property that describes the nature of bare particulars. So it looks as though there can be no such abundant property.

This concern is too quick: the relevant abundant property would correspond to the predicate that holds of a thing just in case that thing has no sparse properties. So the true predicative claim *could* be grounded in a fact about the complex of sparse properties: namely, that no property in this complex is instantiated by the relevant individual. Still there is something puzzling about this idea. Since the particular in question has no sparse properties, there is no

natural/joint-carving/fundamental way in which that particular is. Consider an analogy: it is true that there is no such thing as the golden mountain and that Pegasus does not exist. As such, neither the golden mountain nor Pegasus have any sparse properties. By the foregoing, the golden mountain and Pegasus are bare particulars. But this is a *reductio* for a defense of the thesis that there could actually be bare particulars. We are already committed to the non-existence of Pegasus. So being a bare particular is not existence entailing. That is, the property "being a bare particular" does not satisfy the "being constraint", roughly the principle that anything that is some way thereby exists (Williamson 2013: 148–58). The challenge for the defender of bare particulars, though, was to show that bare particulars could have a nature that consists in failing to have properties. Having a nature that is not existence entailing, though, is not enough to meet this constraint; *ex hypothesi* a nature of this kind is not a way of being.

Let's therefore try a different route. The defender of bare particulars might appeal at this stage to *zero-grounding* (see Fine 2012). (Something is zero-grounded if it is grounded in 0 things.) The empty set, for instance, is plausibly zero-grounded. The abundant property of being a bare particular, we could say, is zero-grounded. There's no need, though, to say something similar about Pegasus or the golden mountain. So we don't run into the problems canvassed in the previous paragraph.

Let's use "P" to denote the (abundant) property of being a bare particular. The claim is now that some particular's being bare is zero-grounded, i.e. that $[Py] \leftarrow []$ for some y. Suppose that the foregoing claim were true. Then the existence of bare particulars would impose no demand on reality. No matter what else is the case, it will be true that []. So we get bare particulars on the cheap. I will now suggest, though, that bare particulars cannot be cheap in this way.

Notice first that bare particulars are *concreta*. Or so I claim. Although I find it intuitive, it is hard to defend this claim in part because it is notoriously difficult to characterize the abstract-concrete distinction precisely (see Lewis 1986: 81–6). Nonetheless the following considerations seem significant: among the hallmarks of *concreta* are their locatedness and particularity. (This shouldn't be understood to mean that all particulars are *concreta* or that all

located things are concrete; the empty set is particular but abstract, space-time points seem to be located, but might be abstract.) For substratum theorists, the *role* of bare particulars is, at least partly, to they serve as grounds for the locatedness and particularity of *thin* particulars. Consider my coffee cup: it has a rich array of properties, including a shape, a mass, and a color. But in themselves these properties are merely abstract, on their own they don't suffice for *this* concrete coffee cup. The idea underpinning substratum theory is that bare particulars serve as "hooks" onto which various properties can be hung (cf. Turner 2011: 51–2). (In Sider's phrase they are "pincushions into which universals may be poked"—Sider 2006: 387.) Doing so allows us to speak of various properties being *coinstantiated* and moreover, being coinstantiated *somewhere*. If bare particulars are to play this role, though, it seems that they themselves must be located, *concrete*, particulars. For it is otherwise hard to see how they could transfer (as it were) their concreteness to the thin particulars they compose.

Sider might disagree. On his picture, numbers, which are paradigmatically abstract, can be thought of as bare members of a distinguished ω -sequence (ibid.: 393). This is a kind of structuralist view. There is, however, something odd about introducing bare particulars into such a picture. We can distinguish between *in re* and *ante rem* forms of structuralism (Shapiro 1997: 85–90). According to the first view (roughly speaking), structures inhere in and are in some sense posterior to the things that realize them. This makes it hard to see how numbers could be bare particulars. For the relevant positions in arithmetical structures are interestingly differentiated; 7 is unlike 4 in that it is prime, 9 and 4 share the property of being squares, etc. Bare particulars though are not differentiated. So we should assume we're dealing with a form of *ante rem* structuralism, on which the structures precede the things that realize them. Still, our view remains in poor shape. Part of the appeal of structuralism is that it may help us to avoid otherwise embarrassing questions of mathematical ontology: why is 2 bare particular *b* rather than bare particular *b*₂ (which is 3)? Swapping *b* and *b*₂ would, after all, make no difference to the way the world is. However, the view under consideration makes such questions salient and indeed pressing. None of this is decisive, but

I think we might reasonably claim that Sider's suggestion that numbers could be bare particulars is less plausible than the contention that bare particulars are concreta.

The remainder of the argument follows swiftly: we do not get concreta on the cheap, their existence *does* impose a demand on reality (cf. Burgess 1998). So the existence of bare particulars must impose a demand on reality. So it cannot be that case that $[Py] \leftarrow []$.

There's a final possibility: perhaps the bareness of some particular needn't be grounded at all, it might be a fundamental fact. The problem with this proposal is that it requires us to use the predicate "P", which recall expresses the property of being bare, in order to articulate how the world is fundamentally. This leads to a nasty dilemma. To see this consider a principle Sider himself defends in other work:

Purity: fundamental truths involve only fundamental notions (Sider 2011: 106)

As Sider uses these terms, we could equally say "fundamental truths involve only joint-carving notions" (ibid.: 5). This principle is highly plausible: the fundamental truths, whatever they are, presumably correspond as far as possible to the basic structure of our world. This in turn seems to require that they are formulated using the joint-carving notions—those that disclose our world's structure. Granted **Purity**, though, the predicate "P" must be fundamental and joint-carving.

At this point, though, it becomes hard to see how P could fail to express a *sparse* property. Sparse properties "carve at the joints...there are just enough of them to characterise things completely and without redundancy" (Lewis 1986: 60). But if so, then P must be among the sparse properties since otherwise the totality of sparse properties would fail to adequately characterize reality. Yet if P is a sparse property, then a contradiction looms: the predicate of being a bare particular, $\lambda x^2.x^2(y^1) \leftrightarrow \neg \exists z^2(z^2(y^1))$, looks inconsistent. The solution was to assume that the value of x is not within the scope of the quantifier here since quantification is understood as implicitly restricted to sparse properties. But if P is a sparse property, then x must be within the scope of the quantifier in which case the predicate is indeed inconsistent. There are, I assume, no true

contradictions, so the friend of bare particulars must reject **Purity**, but this is a substantial price to pay.

3. Conclusion

I suggest, then, that Sider's defense of bare particulars does not succeed. The initial concern was that nothing could be without being *some way*. The proposed solution was to allow that bare particulars do indeed have a nature, since they have an abundant property—namely being bare. But I have argued that there is no satisfactory explanation of how this could be true: the bareness of bare particulars could be a fundamental truth only if we give up **Purity**, which is arguably part-constitutive of what it is for a truth to be fundamental (Sider 2011: 115–16). Therefore, the bareness of bare truths must be grounded in something else. But it can't be zero-grounded since this would allow, counterintuitively, that there could be concreta whose existence made no demand on the world. In that case, the bareness of a particular must be grounded in its not instantiating any sparse properties. That, though, means bare particulars exist in the same sense that the golden mountain exists. Since there are no golden mountains, this amounts to conceding that bare particulars don't exist in any sense at all. Ultimately, then, since there is no way to ground the property of being a bare particular, we should conclude that this isn't a genuine way to be. In that case, though, the original objection to the possibility of bare particulars goes through after all.

References

Burgess, John P. (1998) "Occam's razor and scientific method", in: Matthias Schirn (ed.), *The Philosophy of Mathematics Today*, Oxford: Clarendon Press.

Fine, Kit (2012) "Guide to Ground" In Fabrice Correia and Benjamin Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*, Cambridge: Cambridge University Press.

Lewis, David (1986) *On the Plurality of worlds*, Oxford: Blackwell.

Rosen, Gideon (2010) "Metaphysical Dependence: Grounding and Reduction", in: Bob Hale and Aviv Hoffman (eds.) *Modality: Metaphysics, Logic, and Epistemology*, Oxford: Oxford University Press.

Shapiro, Stewart (1997), *Philosophy of Mathematics: Structure and Ontology*, Oxford: Oxford University Press.

Sider, Theodore (2006) "Bare Particulars", *Philosophical Perspectives* 20, 387–397.

Sider, Theodore (2011) *Writing the Book of the World*, Oxford: Oxford University Press.

Turner, Jason (2011) "Ontological Nihilism", in: Karen Bennett and Dean W. Zimmerman (eds.) *Oxford Studies in Metaphysics* 6, Oxford: Oxford University Press.

Williamson, Timothy (1998) "Bare Possibilia", *Erkenntnis* 48:2/3, 257–273.

Williamson, Timothy (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.

Acknowledgement

I wish to thank Eliya Cohen for her helpful discussion, which was a significant influence on my ideas about this topic. Errors, infelicities, and non-sequiturs are, of course, attributable solely to me.

Why No Digital File Is Unique

Joseph Wang-Kathrein, Ulrich Lobis (Innsbruck, Austria)

Abstract

In everyday life, whenever we work with computer files, we have the feeling that we are working on the same file throughout the time. Just like a notebook, we have the feeling that while the file itself stays the same throughout time, the content of the file may change. However, this (useful) illusion is maintained by the operating system. And this illusion has no technical foundation. In this paper, we want to elucidate why files that persist through time are illusions, how the file system creates the illusion. We will scrutinize possible identity criteria of files. And we will come to the conclusion that every file that shares the same content is identical to each other. In cases where we require uniqueness of files, the only solution is to keep these files in secrecy.

1. We treat files as substances

An author was working on her new novel. On a typical day of work for her, the situation would look like this: In the morning, after she had started her computer, she opened a file with the name “novel.docx”; the file is located in her user folder on the internal drive of her computer. Between 8am and 12pm she would write a few paragraphs and save the progress of her writing a few times. After lunch she spent another three hours on the novel. However, she added only a few lines; mostly she corrected the text she had written in the morning. Before she finished for the day, she saved the file again on the internal drive. Additionally she also copied the file on an external harddisk and renamed the file to “novel20240317.docx” to document her progress and have at least one backup copy.

We assume that people usually create and modify texts in similar ways as the author depicted above. Whether we are writing an article, editing pictures or recording videos, we have the feeling that we are actually working on “files”. Each of the files has its own path and its own name; and we can use word processors or other applications to manipulate the content of the files. Files are like notebooks: Each notebook contains – say – an article we are working on; we write and change the content in the notebook in order to change the content of an article. If we work on several articles at the same time, we must fetch the same notebook each time we want to work on one specific article. Files – at least according to the way we work with them – are similar to these notebooks. Of course, the way we change contents and otherwise interact with

files is very different from notebooks, but we do have to fetch the same file, each time we want to work on one specific article.

And there are other features that make files comparable to notebooks. Files are stored in a certain directory of a certain drive; very much like notebooks are kept in certain compartments of certain shelves. When a shelf is destroyed, notebooks in it may also be destroyed; and when a harddisk has a failure, the files on them may not be readable anymore. We can move a file from one directory to another, just like we can move a notebook from one to another shelf. And we can copy the content of a notebook and put it into a second notebook; similarly, we can duplicate files. Of course, duplicating files is an effortless action; and usually copying one file will result in two files with identical content in it. This kind of perfect copy is very difficult to achieve with “analogue” notebooks.

All in all, in our daily work, files seem to be a kind of Aristotelian substance, they are the “carrier of changes” and persist through time. (Cf. Kanzian/Wang 2008) Of course, when we look into the way how digital drives are managed, we will quickly realize that a file is but a useful illusion that an operating system “creates” for us.

2. Files are Illusions Maintained by Operating Systems

There are many different technologies to persist digital data on a media; and an operating system can use different methods (“file system”) for managing files for each of these technologies. In this paper, we will use the term “file system” to denote the part of the operating system that is used to manage files. Furthermore, we will use harddisks (HDDs) to illustrate what happens when we interact with a digital storage. We believe the basic mode of operation of a file system on HDD is quite comparable to file systems used for other digital media; and what is said here about HDDs can be – mutatis mutandis – applied to other technologies as well.

A disclaimer should be put here: The depiction on the functionality of harddisks presented here is already at a certain “level of abstraction”, i.e. we take up deliberately metaphors and analogies (e.g. “writing zeros and ones”) that are commonly used in textbooks for computer science students.

As is commonly known, on a harddisk information is stored as binary values (zeros and ones) on magnetic discs. These magnet discs are divided into multiple blocks; and the blocks are ordered in such a way that most of the surface of a magnetic disc can be used for storing data. One block is able to store a certain amount of bits (normally in groups, i.e. bytes); and common file systems can attribute one block to one file at most. The attribution of blocks to files is recorded in the metadata of the harddisk.

As harddisks are slow compared to the main memory (RAM), modern computers are programmed to use harddisks as infrequently as possible. The central processing unit (CPU) does not interact directly with the harddisk; it can only work with data already in RAM, and therefore all data needed from the harddisk must be loaded to RAM first. Loading a file from the harddisk, however, is more complicated than simply fetching a notebook from a shelf. When the size of a file exceeds the maximum size a block can keep, the file must use multiple blocks to store its content. And due to the fact that a harddisk does not have unlimited storage space, a file system seldom stores a file in sequentially ordered blocks. It will try to fit all data onto the harddisk as efficiently as possible. When a file is deleted, i.e. when additional free space is available, the file system will try to reuse the freed space for other files. This results in “fragmentation” of files, and the possibility of fragmentation is also the reason why a file must be reconstructed, i.e. the operating system must “rebuild” the content of a file by reading different parts of the harddisk and then putting them back together. (Cf. Tannenbaum 2016, chap. 4; Silberschatz 2019, chap. 13)

Operations on files and directories (e.g. “delete”, “copy” and “move”) are executed in a way that is able to give users the impression that files are persisting artifacts. For example, while we have the feeling that directories are comparable to shelves, they are actually only references to names of other entries in the file system. As we cannot store data in a directory, there are no blocks attributed for a directory on a harddisk. Moving a file from one directory to another, which looks like moving a notebook from one shelf to another, simply results in changing the path of the file, i.e. only one entry in the metadata of the harddisk is altered, while the data on actual sectors of the harddisk isn't changed at all. The path can be considered as part of the name, which is for the OS more part of the metadata than of the data itself. On the

other hand, in order to copy a file to another directory, the operating system must first create a new file and then copy the contents of the blocks attributed to the origin file to other blocks on the harddisk.

It is now clear that files are not something that could exist for themselves on a harddisk. The file system is able to create the illusion of persisting files, but there is really no such thing as persisting files. The file systems are designed in a way that eases our daily interaction with data stored on a harddisk; and the illusion of persisting files makes our interaction with computers easier.

3. When are Files the Same then?

If files are not substances, what are they then? In order to answer this question, we could first ask when are two files the same. As files are “illusions” created by file systems and as we can only interact with the “files” through them, the files cannot be something “outside” of the file system.

Technically, a file system creates the illusion of a file by combining three “parts” of a file: The metadata (including the path and name) of a file, the attributed blocks to this file, and the actual content of the file stored in those blocks.

It seems implausible that just because two files share the same metadata and/or the same blocks on one harddisk, that we call both files “the same”. The reason is simple: Whenever we move a file, we would say that the file stays the same; and after moving a file, the metadata changes. And when we have moved a file to another drive, the file system will use the freed space for other files; this also shows that having the same blocks attributed does not make two files identical to each other.

Can we say that two files are identical to each other, when the contents of the files are the same? One way to investigate whether two files have the same content is to open them using suitable applications and see whether the contents presented on the computer screen (or printed out using a printer) look the same. However, this isn’t sufficient, though. A DOCX file and a PDF file

may share the same textual contents, however, the functionality of both files are very different. A DOCX file contains other metadata than a PDF file; and a PDF file may embed information on used font that is lacking in a DOCX file.

Technically, we could use bitwise identity as criteria for the identity of two file contents. Bitwise identity is the lowest level where we can compare two files. To ascertain that two files are bitwise identical to each other, first, we have to load both files into the main memory of a computer; then we have to compare whether the contents have the same size and whether the zeros and ones at each position of the content are the same. After all these questions are affirmed, we can say that the contents of the two files are the same.

When two files are bitwise identical to each other, applications that are able to use these files will treat them in the same way. We can even construct a digital Theseus Ship example using the notion of bitwise identity: Let us imagine that we have two harddisks (A, B) with the same specification in one computer, while one harddisk (B) is acting as backup of the other harddisk (A). The backup is done as a bitwise copy of the other. One day, as the harddisk A shows signs of degradation, a technician replaces the harddisk A with the backup harddisk (B) and inserts a third harddisk (C) as the backup drive. Now, the B becomes the main harddisk, and C the backup harddisk. As this is done without us noticing it, we have no idea that we are working with different drives now. The question is: Do we work also with different files?

If one accepts that two files are the same if they are bitwise identical, the answer to this question would be “no”. From the perspective of users of the computer, they will not notice that drives have been changed; and neither would they sense that files are changed. From the perspective of the technician, the computer will work in the same way before and after the swapping of harddisks. And even the file system will deliver the same content when they are asked to load data from the harddisk. This shows that the content of a file is actually abstract.

A possible analogy to content of a file could be the memorisation of a folk song by – say – 30 people who are members of a local choir. Let us imagine that the song is memorized in the same way by each person, i.e. there is no difference in text of the memorized poem. Here, it is meaningless to ask whether there

are 30 poems memorized by 30 people – all 30 poems share the same text – or whether there is one poem memorized by 30 people. As long as they sing the same text, there will be no problem for the choir

If this can be accepted, then we can say that – since file contents are abstract – each file can – at least in principle – have multiple instances. And this means that no file is unique in the sense that each substance is unique.

4. Uniqueness through Secrecy?

If what we have said above is true, then we cannot distinguish between copied and original files anymore. However, in our daily life there are situations where we require unique files. This happens e.g. with connected devices on the internet when we need to single out one device from all other devices. Without the ability to single out one computer from others, a “man-in-the-middle attack” (cf. Aakanksha et al. 2019) can be easily performed, as each computer may take over the “identity” of another computer. The common way to counter this kind of attack is to employ special security protocols. Similarly, blockchain technologies (like NFTs) also have the need to single out one specific party from other participating parties. (Cf. Chakravarty/ Sarkar 2020, Part III.)

The common way to achieve the uniqueness is to use cryptographic algorithms to identify and verify servers and parties. To be able to employ these algorithms, certain files (so-called “private keys” and their corresponding “public keys”) must be unique across the internet (resp. the parties in blockchains). A webserver, e.g., makes its public key publicly available and uses the private key to encrypt data it sends out; the receiving client can then use publicly available information (including the public key) to verify that the origin of the data is really coming from a server that has access to the private key – without having the access to the private key itself.

As we have seen, no file can be unique. This means that the only way to have a file with a unique content is therefore keeping these files secret. Private keys may not be shared. And indeed, if the private key is copied to another system, the other system can easily take over the identity of the original system.

References

Aakanksha; Jain, B.; Saxena, D.; Sahni, D.; Sharma, P. (2019) "Analysis of Hypertext Transfer Protocol and Its Variants", in: Panigrahi, B.K. et al. (eds): Smart Innovations in Communication and Computational Sciences, Advances in Intelligent Systems and Computing 670, 171-188. https://doi.org/10.1007/978-981-10-8971-8_17

Chakravarty, S.R.; Sarkar, P. (2020) "An Introduction to Algorithmic Finance, Algorithmic Trading and Blockchain." Emerald Publishing Limited, 2020. <https://doi.org/10.1108/9781789738933>

Kanzian, Ch.; Wang, J. (2008) "Substanzen in der analytischen Ontologie", in: Gutschmidt, H.; Lang-Balestra, A.; Segalerba, G. (eds.): *Substantia - Sic et Non. Eine Geschichte des Substanzbegriffs von der Antike bis zur Gegenwart in Einzelbeiträgen*, Frankfurt, Paris, et. al: Ontos Verlag, 521-542.

Silberschatz, A. et al. (2019) *Operating System Concepts*.

Tanenbaum, Andrew S.; Bos, Herbert (2016) *Moderne Betriebssysteme*. 4., aktualisierte Auflage.

Acknowledgement

A part of this paper will be published as "Illusion der Original-Dateien? Eine kritische Prüfung" in German in "Mitteilungen aus dem Brenner-Archiv" Vol. 42, 2024, 123-141.

A Case for Gender Abolition

Benjamin Weislogel (Columbus, USA)

Abstract

I often receive pushback when I state my position as a gender abolitionist. For a variety of reasons, people often find themselves quite tied to the maintenance of gendered categories for modes of social expression. However, I find the existence of these categories to be quite detrimental to society as a whole, in large part due to the very nature of how categories of gender must be constructed and upheld in the first place. In this paper I outline what I believe constitutes a „social construct“ based upon a conception of rule following informed by later Wittgenstein’s work, defining distinctions between certain kinds of rules and how they function, for the purposes of determining the forms which social constructions might take. From this, I provide an account of what kinds of rules I believe compose categories of gendered modes of expression and how they are assigned, applied, and enforced. I then argue that the nature of these rules necessitates that gender exists as a superficial social category which is historically used to enforce harmful social roles, and that the creation of new non-binary categories of expression is not sufficient to resolve these inherent issues, and that abolition of the application of gendered categories to human modes of expression is the only route by which these harms can be mitigated.

As a non-binary person and self-described gender abolitionist, I find that during conversations about the nature of gender I often explain or justify my view by stating something akin to, “Well, gender is a social construct.” While this statement is undeniably true, as there are many people who have made this argument in the past, my reference to this may strike some as odd, as it has myself at times. What does its being socially constructed say about the nature of gender, really? We can’t say that social constructs are necessarily bad or harmful, nor that they are unnecessary or unnatural, nor that it justifies any implication that it is something worth abolishing. There are plenty of social constructs for which some or all of these are untrue, language, trade, family, or law, among others. All being a social construct seems to tell us is that it is a product of society, that it is something enacted and enforced by social means, and that it is formed by rules intended to govern certain types of interactions among people. To discern the nature of gender, and how it might be harmful, perhaps an account of what a social construct is, in the first place, is called for.

It is fairly obvious by the name that these are something which is constructed by people, within and by way of the existence of a society. What constitutes a society may vary, but simply put it is a complex web of interpersonal relations,

a complex web of interaction between people which is governed by some system of rules. I am not certain this is a common view, but it would seem to me that rules are a necessary component of social complexity, whether one upholds a more Hobbesian view of society, or a more cooperative one.

I posit the following conception of rules based upon much of what Wittgenstein himself has said about them, including my interpretation of the rule following paradox (PI 2009: §201). First and foremost, rules do not exist to be strictly followed, nor can they be. Instead, they act as a blueprint, a set of instructions by which a particular outcome might be achieved. Rules do not exist without reason. Rules are outcomes oriented, that is, adherence is driven by the outcomes they point us towards, and their success relies on how well they achieve these outcomes. In this sense, they are, by and large, arbitrary. The only thing a rule can be judged by, is how effective it is. The method it provides for achieving its goal can often be construed in many different ways. Any given rule might be considered good or bad based upon a multitude of factors, such as how it is implemented, what it achieves, and how it is enforced, but simply being a rule does not grant something any sense of goodness, nor of effectiveness. These are determined by what its implementation results in, whether those results align with desired outcomes, and whether the outcomes can be said to be beneficial for society generally. The rule itself says nothing about its outcomes, only guides us to one.

In the case of social rules, their aim is to maintain a given social structure, whatever form that may take. This is accomplished by limitations placed upon what we consider to be socially acceptable behavior. When making determinations about something with as broad a scope as the whole of possible human behavior it is not particularly useful to build this upon positive rules. What I call positive rules are ones which obligate a certain type of action, that is, you are expected to act in a certain way, in accordance with it, while a negative rule is one which restricts actions, dictating what we are expected not to do. Positive rules, for society generally, are bound to be excessively broad and generalized. These are manifest in phrases such as “Be kind”, or “Respect your elders”. Even something such as “Pay your taxes” doesn’t possess much in the way of meaningful content. They are not very useful unless we know what specific sorts of actions constitute these ideas, such as “kindness” or “respect”, and the actions which fall into these categories

vary widely among cultures and traditions. It is negative rules, then, that are far more useful, and thus more commonly enforced in social contexts, so it is by way of these sorts of rules that social constructs are, well, constructed.

Maintaining social cohesion is vitally important because, given a large enough group of people and a complex enough society, there is no avoiding social conflict. No two people are going to have perfectly aligned interests, no matter how agreeable they might be. Social conflict is an inevitability, and it is disruptive. It threatens the general social order. Social constructions exist to help manage resolution of social conflict so that social order can be maintained. If a social conflict grows to a large enough scale, there is often a significant amount of social upheaval, as people tend to take sides in a given conflict and the old social structures break down, while new ones, upheld by either side of the conflict according to their desires, are brought about. Again, none of these things is necessarily good nor bad on its face. Their value is determined by how these constructs are enforced, and what outcomes they lead to.

There is a distinction to be made between social rules and interactive rules. I made mention earlier that language falls into this category of social constructions, but it is not wholly governed by social rules. Language is a tool that allows for social interactions to take place. It is a game we play amongst one another. The rules of language are just as much outcome oriented as any other, but in a primitive sense language is governed by positive rules, rather than negative ones. This is because language is constructed from the ground up and require us to know what the initial “rules of play” are. If I need to determine how I convey to the person in front of me that I want to purchase a dozen eggs from them, I need to know what sorts of actions count for conveying this information before I can place any restrictions on the interaction. Once I have a system for determining what an effective move in the game is, then I can restrict the effective moves to a select subset of moves which might be considered “respectful” or “kind” or the like, by imposing restrictions such as not using profanity to refer to them or their stock. The rules which tell us what words or phrases we shouldn’t use in certain contexts are not linguistic in nature, rather they are social rules which place restrictions upon the linguistic game.

There is a further distinction to be made regarding types of social rules. I will call the first “explicit” and the second “implicit”. Explicit social rules are those which are clearly laid out by an authoritative governing body, which enforces them via institutional power. These are rules conveyed by way of institutional decree, such as laws, regulations, proclamations, or ordinances. On the other hand, implicit rules are enforced on the level of social interactions by way of the interpersonal relationships we all share with other people in society. Implicit rules are largely enforced socially as well, most often with certain linguistic patterns that convey the social impact of their violations, such as disappointment, shame, irritation, or disdain.

So, what do we make of social constructions such as gender, race, sexuality, and the like? These are all defined by sets of social rules which exist to maintain a certain social organization between various categories of people. This is accomplished by placing restrictions on the actions and expressions that are considered acceptable for people to engage in based upon the respective group they have been determined to belong to.

Gender in particular is defined most strictly as consisting of two categories, man and woman. There have been many attempts to positively define what a given gender is, including reactionary tendencies to appeal to chromosomal composition, primary sex characteristics, or the like, despite most often relying on secondary sex characteristics to make judgements about a person’s assumed category of gender assignment. More critical approaches have been taken, such as Haslanger’s well known work defining women in relation to systemic oppression (Haslanger 2000), or Jenkins’ response and appeal to identity for the sake of the inclusion of trans women (Jenkins 2016), as well as contemporary approaches which seek to define categories that exist outside the gender binary (Dembroff 2020).

I find most accounts insufficient for constituting generalized gender categories or identifying some unique necessary property shared by all who are assigned to one. Instead, the nature of social rules and how they are implemented leads us to construct gender categories that aren’t defined by a common feature, but by what an adherent should not be, by appealing to rules which govern what highly variable boundaries exist for any given gender category. What a “woman” is, is whoever a society accepts as fulfilling its criteria of

womanhood, and those criteria vary even within a given society. These criteria constitute this category as being defined in a manner which Wittgenstein has termed “family resemblances” (PI 2009: §67). There simply is no one way to be a woman, but there are ways one can be not enough of one, criteria which some consider to justify exclusion, and the same is true of traditional conceptions of manhood.

This sets gender apart from other categories such as race or sexuality, where one can appeal in some respect to clearly observable phenomenon. Categorizations of race and ethnicity are based upon skin tone and national origin, ancestry, and cultural background. Sexuality is based largely on who a person is attracted to. This is self-reported, but directly observed by the person themselves. There is a concrete basis for its foundation. Gender, however, overlies no such clearly defined property of a person’s being. It is at once a social role governed by expectations about one’s actions, and something one expresses, determined by the interaction between one’s identified mode of expression and the one to which they are assigned by the public. For those who adhere to their assigned gender, they correct their behaviour to fall in line with certain, although not necessarily all, parameters of expected behavior. They may define themselves based around looser or stricter rules, but they exist within a space which has been deemed an “acceptable way” to be part of that category. For binary trans people I have often observed that they trend toward trying to adhere to some apparent nucleus of gendered expression, existing within much stricter parameters for their desired mode of expression in order to justify to society that they fall in a different category than what was originally expected of them.

However, in order to adhere to such a category the boundaries of it must be defined to us. The social interactions which do this most often take the form of “corrective admonishments”. They are often linguistic expressions which are recited whenever a “violation” of some rule has been observed. Consider the more general social rules such as ‘Men should not appear weak’, or ‘Men should not express negative emotions that are not anger’. People who enforce traditional gender roles, whether consciously or not, will often recite phrases such as “Boys don’t cry”, “Be a man”, “Suck it up”, “Grow a pair”, or “Stand up for yourself”, in response to violations of the above rules. These phrases tend to be applied to more minor violations as a means of defining more specific

rules. For more “severe” violations, far more volatility may be employed in order to “correct” the observed deviation from what is deemed “acceptable” social conduct. We see this often expressed as mockery and condemnation which may escalate further into acts of violence.

Notably, we can observe within the rules used to defining the boundaries of womanhood, the enforcement of a hierarchical power dynamic. Rules employed to define acceptable conduct for women are often defined by their relationship to men, such as not talking over or interrupting men, not talking back, undermining a man’s authority, not being loud, aggressive, or contrarian, not being opinionated or argumentative. Such rules regulate a woman’s behavior by how she is expected to interact with men, whereas rules restricting men’s behavior often create restrictions independently of any appeal to the potential presence of women. Because of this, men tend to adhere to a set of general unspoken principles, which are the criteria for “manliness”, while women are expected to minimize their presence and impact they might have on the men around them, which become the criteria for “womanhood”.

In more recent years we have seen an uptick in gender non-conformity and non-binary identities. These terms refer to people who exist in ways which actively and consciously defy the social rules which define traditional gender categories. While new categories absolutely grant people a wider variety of expression, I am not convinced that the issues inherent in socially constructed notions of gender can be resolved by their presence. Even though non-binary identification is supposed to allow one to exist outside the boundaries of gender categories, we often see these people, myself included, continuing to be automatically associated with whatever their perceived traditional category is by society at large. These identities are often treated as invalid on the basis of the person expressing themselves in a manner too heavily associated with a binary category, not seeming to be “androgynous enough”, or being superficially acknowledged while the underlying habits of social category enforcement continue to be applied to them. Further, with the advent of novel categories that exist in the non-binary “space” which seek to define themselves as a strict category in their own right, we see new social rules continuously implemented in order to police behaviors of those who seek to identify with those categories.

While a category such as race can be disarmed to some degree by the admission that they are defined by little more than skin tone, gender is something which is constructed solely as a set of social rules which enforce vaguely defined categories based upon little more than instinctual assumptions of ones outward appearance. Gender is, by necessity, a set of highly restrictive categories which are used to justify negative, harmful, and even violent reactions toward non-conforming behavior, and which are not a necessary component of self-expression. They define themselves by a swath of acceptable and unacceptable actions and modes of expression, which can be engaged in whether they have been associated with a gender category by society at large or not. Decoupling gender categories from their range of acceptable actions is not possible in its entirety without making the distinctions fundamentally meaningless, but so long as these associations exist, people will continue to have their behavior and modes of expression policed by society based upon superficial assumptions made at a glance. It is for these reasons, among others, that I remain a gender abolitionist.

References

Dembroff, Robin (2020) “Beyond Binary: Genderqueer as Critical Gender Kind”, *Philosophers’ Imprint* 20 (9), 1-23.

Haslanger, Sally (2000) “Gender and race: (What) are they? (What) do we want them to be?”, *Noûs* 34 (1), 31–55.

Jenkins, Katharine (2016) “Amelioration and Inclusion: Gender Identity and the Concept of Woman”, *Ethics* 126 (2), 394-421.

Wittgenstein, Ludwig (2009) *Philosophical Investigations / Philosophische Untersuchungen*. Ed. by P. M. S. Hacker and Joachim Schulte, translated by G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. New York: Wiley.

Can Wittgensteinian Use of Language Explain (Away) Intentionality and Mental Representation?

Christian Helmut Wenzel (Taipei, Taiwan)

Abstract

In the *Tractatus*, Wittgenstein uses “*Darstellung*,” but not “*Vorstellung*.” Both words are usually translated into English as “representation,” which causes confusion. The former is used in mathematics, but only the latter means something mental. Wittgenstein talks of “thought” (*Denken*), but assimilates it to his picture theory, which again is not about the mental. In the *Philosophical Investigations*, he does talk about *Vorstellung*, together with expecting, wishing, thinking, and other psychological phenomena. But he merely observes how the expressions are used. He does not use them himself. He is skeptical about psychology, also in connection with brain science. Instead, he emphasizes outer expression, especially in language. But I think this comes too late, because intentionality and mental representation already occur in perception, which is more basic and does not require language or other forms of public expression. Thus, I think Wittgenstein does not explain intentionality and mental representation. Nor does he explain them away. I think Aristotle’s *De Anima* and the discussions in the Middle Ages about *esse intentionale*, *repraesentativum*, *obiectivum*, *cognitum* and *apparens* were on the better track, when many of our key terms we use today were coined.

1. Pictures in the *Tractatus*

Did Wittgenstein in the *Tractatus* deal with intentionality and mental representation? Hardly. Scott Soames’ recent book *The Analytic Tradition in Philosophy*, vol. 2, *A New Vision* has a long chapter on the *Tractatus* (pp. 3-103), but the words “representation,” “mental,” and “intentionality” are not to be found in the index.

The *Tractatus* deals with “the world,” “facts,” and “propositions.” It is some kind of minimalist transcendental philosophy. But it does not deal with consciousness and judgments, at least not in Kant’s sense of “judgment.” Remarkably, analytic philosophy in general does not think much about “judgment,” whereas classical German philosophy did about “*Urteil*” (Wenzel 2011). An *Urteil* requires a subject that makes it, a proposition does not. That is the major difference. The *Tractatus* belongs to that analytic philosophy.

Besides an early interest in Schopenhauer, I do not think Wittgenstein was familiar with classical German philosophy. He was more influenced by Frege, for whom thoughts are in a “third realm.” The question of how we “grasp” them, Frege left to psychology. Frege, avoided psychology, and Wittgenstein

followed him in this. I think this lasted throughout his life. But I also think the he nevertheless struggled with questions about the soul in ways that I think are not easy to explain. But this is another topic, touching on ethics, aesthetics, and religion (Wenzel 2010).

In the *Tractatus*, Wittgenstein speaks of “pictures” that we “make for ourselves” (*Wir machen uns Bilder der Tatsachen*, 2.1, translations from Wittgenstein are always my own). But he does not further investigate the “making” and the “for ourselves,” nor how we “have” pictures. Instead, there is an abstract “correspondence” between “elements” in the picture and “objects” in the world (2.13). Importantly, the picture too is a fact (2.141), and since the world is the “totality” of facts (1.1), the picture too belongs to the word.

There seems to be no room for the perceiving, feeling, and thinking subject in the *Tractatus*. The subject shrinks to a point (5.6331). It is not part of the world but its “limit” (5.632). Wittgenstein briefly entertains the idea of solipsism (5.62), most likely influenced by Schopenhauer, but does not develop it. Solipsism too shrinks to a point, but is still said to be real (5.64) (Wenzel 2023).

What is merely a point or a limit, cannot be a mind that is complex enough to have *Vorstellungen* and intentionality. Kant did not conceive of the thinking subject as a substance either, but he attributed a rich faculty of cognition to it that he understood as the source of *a priori* conditions of the possibility of cognition. Wittgenstein dismisses such an idea, saying without further argument that “there is no *a priori* order of things” (5.634). If there is an *a priori* logic in the *Tractatus*, it cannot not be Kantian transcendental logic, which would involve intuition, synthesis, categories, and consciousness. In the *Tractatus*, here is no room for some kind of psychology of mental representations.

Importantly, when investigating the relationship between picture and object, Wittgenstein talks of “*Darstellung*,” but not of “*Vorstellung*” (2.173, 2.174, 4.242). In the English translation, the difference is lost. Usually both German expressions are translated as “representation.” But they mean different things (Wenzel 2021, 15-16; Wenzel 2023, 633-634). A sketch on a piece of paper would be called a “*Darstellung*,” but not “*Vorstellung*.” Wittgenstein talks of *Darstellung* in relation to pictures and signs. They are not something mental.

He talks of propositions, musical scores, and phonetic spelling that “represent” (*darstellen*, 4.011). In German, one could not say “*vorstellen*” in these contexts.

Kant extensively uses the term “*Vorstellung*.” Importantly, a *Vorstellung* requires a subject that has and entertains the *Vorstellung*. It needs a bearer (unless one talks of a theater performance, which is also called a “*Vorstellung*”). In German, one says “*Ich mache mir eine Vorstellung*.” Translated literally into English, it would be “I make myself a mental representation.” But this is not ordinary English.

A *Vorstellung* that I make myself is not something physical that can be shown to others and shared in that way. Again, a *Vorstellung* is something mental. My *Vorstellung* is not yours, similar to my pain. Wittgenstein avoids talk of the mental. The word “*Vorstellung*” is one of the central terms in Kant’s first *Critique*, arguably *the* most basic. But it never occurs in the *Tractatus*.

Wittgenstein in the *Tractatus* mentions “thought” (3), but then incorporates it into his theory of “pictures” (3.01), “sentences” (3.1), and signs (3.12; 3.5). He uses the metaphor of a “method of projection” (*Projektionsmethode*) to try to cast some light on the idea of “thinking the meaning of a sentence” (*Das Denken des Satz-Sinnes*, 3.11). But this does not offer much of an explanation (Amereller 2001). There is simply no room for the thinking subject that could possibly have mental representations (*Vorstellungen*). At best there are *Darstellungen*, but these are not *Vorstellungen*.

2. Use of language in the Philosophical Investigations

The *Philosophical Investigations* are very different from the *Tractatus*. They are more fine-tuned, pay attention to individual cases, and avoid generalization and abstract theorizing. Differently from the *Tractatus*, in the *Philosophical Investigations*, Wittgenstein indeed does talk about *Vorstellungen*, or rather about the word “*Vorstellungen*.” He does not trust *Vorstellungen*. He prefers not to use the word himself, but to observe how others use it. He keeps his distance. This happens roughly from §370 to §460, when he talks about expecting, wishing, planning, thinking, and similar psychological phenomena. They are difficult to observe from the outside.

From the outside, expecting someone to come is one thing, someone coming is another. They look very different, he says. He asks how expectation and fulfillment can possibly meet, and how one possibly could compare them. But similar to the *Tractatus*, he falls back on language. He says that the sentence “I expect he is coming” can be seen to meet with the sentence “He is coming” (§444). Thus,

It is in language that expectation and fulfillment touch each other. (§445)

He argues that the fulfillment cannot possibly be in the expectation itself. Similarly, the bang of a gun for example that I expect cannot be in the expectation (§442), and the same holds for other mental phenomena, such as wishing for something that does not exist, planning what has not yet been done, thinking about things that are not real. This is reminiscent of the *Tractarian* metaphor of the elements of a picture being “feelers” that “reach out” and “touch” reality (2.1515).

Wittgenstein does not trust introspection and the first-person perspective generally. His approach is the opposite of Husserl’s phenomenology, who wants to observe our mental states from the first-person point of view and who is ready to bracket the external world.

Husserl would explain both the expectation and the fulfillment from the first-person perspective. That can be done. But Wittgenstein prefers to look at things from outside. He wants outer criteria. He holds on to language and use. This applies also to *Vorstellungen*. Thus, he writes:

One ought to ask, not what *Vorstellungen* [mental representations, acts of imagination] are or what happens if one imagines something [*sich etwas vorstellt*], but how the word “*Vorstellung*” [mental representation, imagination, act of imagination] is used. (§370)

Anscombe does not translate “*Vorstellung*” as “representation.” She uses the word “image.” This indeed is better. But a *Vorstellung* is something mental, while an image does not need to be. The German word “*Vorstellung*” is difficult to translate into English.

If one wants to avoid introspection and the first-person point of view, one might want to turn to cognitive science, neuroscience, and brain research. This

would be a peculiar way of looking outside-in, and it can be done, too. But Wittgenstein is not impressed by this idea either. In the *Remarks about the Philosophy of Psychology*, we find many remarks that reject this possibility.

No supposition seems to me more natural than that there is no process in the brain correlated [*zugeordnet*] with associating or with thinking; so that it would be impossible to read off thought-processes from brain processes. I mean this: if I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the *system* continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos? (RPP I 903)

He prefers to observe outer behavior and the use of language (outside-in) over introspection and the first-person point of view (inside-out) as well as attempts to include neuroscience (outside-in). For a critical discussion of the RPP passage, see Wenzel 2022 and 2019. Even God would need outer criteria, he says.

If God had looked into our minds [*unsere Seelen*] he would not have been able to see there who we were speaking of. (PI II: 221)

I will counter this by relying, somewhat ironically, on meaning externalism.

3. Mental representations are real and can exist without language: contra Wittgenstein

Tim Crane (2014) has already argued that we should reject Wittgenstein's idea that the relationship between thought and reality can be explained linguistically, saying for instance that expectation and fulfillment meet in language (§445). Contrary to what Hacker claims, Crane argues that this is not a "resolution of the problems and puzzles of intentionality" (Hacker 1996: 129). I very much agree. But Crane completely misses the distinction between "*Darstellung*" and "*Vorstellung*" when quoting Wittgenstein. He always writes "representation," conflating the two, which distorts the problem. In the *Tractatus*, Wittgenstein uses "*Darstellung*" (and very often "*darstellen*")

himself, without criticizing it. He never mentions “*vorstellen*” or “*Vorstellung*.” In the *Philosophical Investigations*, he writes about “*vorstellen*.” But he does not use it. He criticizes its use by others, and tries to dissolve it.

Jay Garfield (unpublished) criticizes what he sees as a too widespread use of the word “representation,” for instance in neuroscience and cognitive science by Churchland and Clark. It is a misleading metaphor, he argues. He says that we can learn from Wittgenstein that “representation is essentially a public, convention-governed phenomenon” and that “representation is symbolic” (Garfield, unpublished). There is intentionality without representation, he argues. I disagree. Representation does not need to be public and symbolic. It is more basic. Garfield does not pay attention to the original meaning of “*Vorstellung*,” and like Crane, he does not distinguish between “*darstellen*” and “*vorstellen*.”

I think we should notice that perceiving, imagining, and dreaming involve ways of *vorstellen*. Pre-linguistic animals do that. They have *Vorstellungen*. In Kant, intuitions (*Anschauungen*) are *Vorstellungen*. They are involved in perception. This of course creates the problem of non-conceptual perception in Kant. It raises the question whether we share perception with animals. But that is a further issue (Wenzel 2005).

In general, historically and systematically, I think we should take seriously the phenomena of imagining and dreaming that gave rise to the problem of “mental representation” in the Middle Ages, when many of the terms that we still use today were coined, such as *esse intentionale*, *intelligibile*, *repraesentativum*, *obiectivum*, *cognitum*, and *apparens*. Philosophers in the Middle Ages asked what it is that we see in dreams and imagination. Thinking about these phenomena, when no real and mind-independent object is present, we realize that in ordinary perception too, we do not see things directly and in themselves. Perception always involves imagination. All we have are appearances that we and our minds-brains create, sometimes with and sometimes without input from outside.

We should of course not “freeze” these appearances by thinking of them as mind-independent objects. Brentano and Kant knew that. They understood *Vorstellung* as the act of *Vorstellen*. Mental representations (*Vorstellungen*) are

not things but acts, and as such they are real. The problem was much discussed by *Malebranche* and *Arnauld* soon after *Descartes* and his use of Latin *idea* and French *idée*. Descartes was not clear about the act-object distinction. These are the roots of the problem of mental representation.

Intentionality and mental representation, whatever it is, begins with perception and involves consciousness. Pre-linguistic animals have it. Wittgenstein's focus on sentences comes too late, even if he pays attention to how sentences are used. Scholars in the Middle Ages very much focused on Aristotle's *De Anima*, which is about perception, not language. Brentano was very familiar with Aristotle when he introduced the term "intentionality." He writes of *Vorstellung* in Kant's sense of the word, thinking of acts and perception. Mental representation does not need to be symbolic. It is more fundamental.

One might think that Putnam's meaning externalism supports Wittgenstein's demand for external criteria. But when we imagine how to arrange pieces of furniture, or when we do research in mathematics and think in pictures, there are no external objects causing our acts of imagination and visual thought. We do it all from within, relying on memory. According to meaning externalism, we then should say that we see the causes, which in this case are neural events in the brain, maybe in connection with our whole body! The causes are not external objects but inside ourselves. Of course, there is memory and there were causes outside in the past. But memory keeps working on itself. All this would force Wittgenstein to consider neuroscience and the brain, which he avoids, as he similarly avoids first-person talk of *Vorstellungen* and mental phenomena general. He avoids *both* the brain and the mind.

One might say that Putnam's meaning externalism is about words and not perception. But words require images, as already Aristotle pointed out.

Thus, ironically, meaning externalism does not help. It would lead to internalism, not regarding the mind, but regarding the brain, another kind of "inside." Not only in dreams and pure imagination but also in ordinary perception, the brain creates the appearances and impressions that we have. Perception is not passive. It is active and involves imagination. We of course say that we see people and mountains, not their appearances. But all we

actually “have” are appearances. Wittgenstein does not like that use of “have” in the “having” of *Vorstellungen*. But he cannot explain it away by focusing on sentences. In acts of perception, we create appearances with the help of input from outside. But in dreams, day dreams, and imagination, we do it without that external help. Philosophers in the Middle Ages took that seriously.

Wittgenstein does not like the use of “subject”, “I”, and “to have” in “to have *Vorstellungen*” (§398, §402). He wants outer criteria and looks for them in public behavior and ordinary use of language. Today, fMRI is another way of providing outer criteria. Admittedly, these are not ordinary outer criteria, and they still depend on whoever “reads” fMRI and understands what he or she is looking for (Wenzel 2022 and 2013). The idea that the appearances we have are created by the brain, was known in the Middle Ages. One does not need to know about fMRI to arrive at that idea. From there, arguments for indirect realism and idealism naturally arose. Something similar happened one thousand years ago in Yogācāra Buddhism in India with Vasubandhu, Dignāga and Dharmakīrti.

The first-person point of view as well as brain science cannot be avoided by looking at language and the ways sentences are used as Wittgenstein suggests. Language is rooted in perception, which has wider scope than language and is not public. Sometimes we think in pictures. Thought can be below the level of awareness, subconscious and unconscious. It can be chaotic, flimsy, vague, or fine-grained. Sometimes we do not have words for it. Language is only the tip of an iceberg (Wenzel 2022). Thus, language and its use come too late. They cannot fully explain intentionality and mental representation. Aristotle’s *De Anima* was on a better track.

References

- Ammereller, Erich (2001) "Die abbildende Beziehung. Zum Problem der Intentionalität im *Tractatus*", in: *Ludwig Wittgenstein: Tractatus Logico Philosophicus*, Willhelm Vossenkuhl (ed.), Akademie Verlag, 111-140.
- Crane, Tim (2014) "Wittgenstein on Intentionality (2010)", in: Tim Crane, *Aspects of Psychologism*. Cambridge University, 40-60.
- Garfield, Jay (unpublished draft) "Intention (Doing Away with Mental Representation)."
- Hacker, P.M.S. (1996) *Wittgenstein's Place in Twentieth Century Philosophy*, Blackwell.

Soames, Scott (2018) *The Analytic Tradition in Philosophy*, vol. 2, *A New Vision*. Princeton University Press.

Wenzel, Christian Helmut (2023) "Solipsism and how to read Wittgenstein's *Tractatus* 5.62", *100 Years of Tractatus Logico-Philosophicus. Contributions to the 44th International Wittgenstein Symposium*, vol 28. Kirchberg am Wechsel, 631-638.

Wenzel, Christian Helmut (2022) "Can Thoughts be Read from the Brain? Neuroscience contra Wittgenstein", *Synthese* 200:183.

Wenzel, Christian Helmut (2021) "How Representational is the Mind? Introduction and Overview", *Philosophy East and West* 71/1, 13-37.

Wenzel, Christian Helmut (2019) "Wittgenstein über Gedanken und Gehirnprozesse", in: *Crisis and Critique: Philosophical Analysis and Current Events. Papers of the 42nd International Wittgenstein Symposium*, vol. 27, Kirchberg am Wechsel, 267-269.

Wenzel, Christian Helmut (2013) "Does Thought Happen In The Brain?", in: *Mind, Language and Action. Papers of the 36th International Wittgenstein Symposium*, vol. 21, Kirchberg am Wechsel, 453-455.

Wenzel, Christian Helmut (2011) "Urteil", in: Petra Kolmer and Armin Wildfeuer (eds.), *Neues Handbuch Philosophischer Grundbegriffe* (Encyclopedia of Fundamental Concepts in Philosophy), 3 vols., Verlag Karl Alber, 2284-2296.

Wenzel, Christian Helmut (2010) "On Wittgenstein's Notion of Meaning Blindness: Its Subjective, Objective, and Aesthetic Aspects," *Philosophical Investigations*, 201-219.

Wenzel, Christian Helmut (2005) „Spielen nach Kant die Kategorien schon bei der Wahrnehmung eine Rolle? Peter Rohs und John McDowell“ (Do the Categories According to Kant Play a Role Already in Perception?), *Kant-Studien* 96/4, 407-426.

Wittgenstein, Ludwig (1958) *Philosophical Investigations*, translated by G.E.M. Anscombe. Second edition, Blackwell.

Wittgenstein (1980) *Remarks on The Philosophy of Psychology (RPP)*, two volumes, translated by G.E.M. Anscombe, Blackwell.

Overcoming McDowell's Dualism of First and Second Nature

Yara Windmüller (Kiel, Germany)

Abstract

John McDowell's philosophy is well-known for tackling dualisms concerned with the relation between the individual and the world. One question that he leaves open, however, is how the dualism of first and second nature can be overcome. The paper is devoted to sketching a way towards answering this question by taking the Wittgensteinian and McDowellian therapeutic approach seriously also where human nature is concerned. It shows that it is not enough to mediate two conceptions of nature; instead, this is just a necessary step on the way of climbing the full metaphorical ladder. I claim that the endeavor is only successful where again just one conception of nature is put in place which entails all aspects that current debates on first and second nature highlight.

1. Introduction

Many dualisms that the history of philosophy has introduced to us seem hard to overcome today. Stitching both sides back together is often impossible, and it often seems that there is no satisfactory way to calm such trouble. John McDowell's philosophy is well-known for tackling such dualisms concerned with the relation between the individual and the world, and he does so in a typical Wittgensteinian manner: McDowell shows that many dualisms turn out to be no more than the result of misleading theories about how those matters are related, and he usually offers a way out of such inaccurate conceptions. In doing so, he does not take a stance somewhere within the field of such dualisms, but rather shows how to overcome them by undermining the preconditions under which these dualisms come into being in the first place. The approach is a thoroughly therapeutic one, which patiently discusses the apparent options that philosophy suggests before showing that all of them rest on some intolerable foundation that leaves us unsatisfied in some way. McDowell calms these worries that arise from the lack of satisfaction by offering a new option—one which usually identifies the aspect that leaves us worried before showing that we can view things a little different to overcome such issues.

In this paper, I focus on the common distinction between first and second nature, which McDowell introduces in a way that makes it dualistic: He discusses both natures as mutually limiting, whilst each nature has governance over some space that remains isolated from the authority of the

other. This way, quite uncharacteristically for his philosophy, McDowell does *not* overcome this dualism but rather reinforces it by building upon the dualistic arrangement of these natures (section 2). I suggest that this is a rare inconsistency in his philosophy, and I aim to overcome it by picking up the very same McDowellian argument and developing it further (section 3).

2. Diagnosis

An essential part of McDowell's discussion of how it is possible for us as rational beings to get access to the world includes dealing with human nature, the very issue that opens up the dualistic picture: The first way to characterize nature falls under the label of *first nature*, describing a dominant assumption about nature which McDowell criticizes throughout his book. First nature gives credit to natural sciences, which are often said to provide us with all the facts and all the laws we need in order to fully explain a human's being in the world. They tell us that gravity exists and that there are certain biological facts that define the things we can do: We need our legs and certain muscles to be able to walk, and unlike birds we are unable to fly because we lack wings (de Gaynesford 2004: 57). The natural sciences also delineate the boundaries between our bodies and the world around us, and they make efforts to discover the functioning of our brains and their goings-on. The histories of philosophy and science make us conceive these discoveries as facts about us and the world which we are subjected to without having much room to avoid their force. In this sense, first nature entails no normativity or meaning (Testa 2007: 483): it just plainly puts limits to the exercise of our abilities (McDowell 1996: 109; McDowell 1998: 190; see also Haddock 2009: 65).

However, there are matters that first nature cannot explain, but which still seem quite substantial to our daily lives. This is why, drawing on Aristotle, McDowell reminds his readers of a characterization of nature as *second nature*. Second nature entails, broadly speaking, the aspects of life that rely on the acquisition of *practical wisdom*. Practical wisdom gives a person reasons for acting in certain ways, which—unlike first nature does—make use of considerations of normativity and ethics all the way out (McDowell 1996: 78-84). This way, second nature entails more than just plain scientific facts, and in a way it gives us—those who are equipped with second nature—

capacities to reflect upon and judge our acting. These capacities are acquired by every human being throughout the process of their upbringing. Already maneuvering within second nature, years of practice slowly equip children with abilities to critically evaluate reasons and motivations for acting, some of which children develop all by themselves. Other reasons for acting are already collectively agreed upon, and children's eyes are being opened to them over the course of their upbringing (Rapp 2014: 162). This development is part of the very same process that makes children learn a language and that introduces them to a tradition. This also implies that all these second-natural goings-on are both already conceptually structured and permeated with normative considerations (Bertram 2014: 122-124). This marks the clear distinction of second to first nature, which—as I noted above—is merely open to facts, leaving no room for such normative statements.

Now it would be possible to conceptualize first and second nature as two totally separated spaces, creating a strict dualism with no way to mediate the two. Of course this is no satisfactory choice, and it is also not what McDowell endorses. Instead, first and second nature stand in relation to each other as they depend on each other in the very process of making sense of the world. Second nature equips us with capacities of ethical thinking and practical wisdom, which entail the standing obligation to reflect on our acting. We can only do so not from sideways-on, but only *from within*, i.e. from within an already existing practice of second nature, which requires exactly this sort of ethical upbringing that children receive (McDowell 1996: 81). Even though this is the prerequisite for any rational acting, the capacities of second nature do not flow free of any constraint. First nature is responsible for providing these constraints; it must, in McDowell's words, “put limits on the shapings of second nature that are possible for [human beings]” (McDowell 1998; 190). The existence of such limitation is crucial, because otherwise we would run the risk of losing the very grip on the world that we constitutively need to prevent us from falling into coherentism, a conception we want to avoid. We always need both first and second nature in order to explain our being in the world—second nature to explain the ways we rationally make sense of what is around us, and first nature to present us with certain options from which we can choose. In this sense, second nature is often focused on when discussing these matters, whereas first nature really seems to be regarded as providing not

much more than such a limiting function (see for example Müller 2017a: 306; Müller 2017b: 160; Testa 2007: 486).

This means that—even though we do deliberate upon certain aspects of nature and their belonging to either of the two spaces—some distinction always remains. It is important to note that it is not only a terminological distinction that McDowell endorses to separate certain aspects of what is actually the same thing. Instead, he explicitly rests his argument on the separation of the two natures, which gets spelled out where first and second nature are conceptualized as mutually delimitating. Unfortunately, the upholding of this distinction leads to a dualistic picture of first and second nature, which McDowell himself does not tackle. In other words: If McDowell were consistent in his treatment of dualisms, he should be attempting to propose a therapeutic way out of the dualism of first and second nature as well, instead of building his endorsement of second nature based on the limitations of and by first nature. However, though he discusses second nature and its priority status in quite some length, it seems that he does not go as far as letting go of first nature completely. Occasionally throughout such discussions he states that “[o]f course first nature matters” (McDowell 1998, 190), or he calls the contents of the space of nature “natural in the relevant sense” and our “animal nature” (McDowell 1996: 74). Such statements reveal a falling back into exactly the way of thinking that preserves the distinguished status of first-natural terminology, which effectively pushes us back to where we started.

3. Therapy

In short, McDowell's lack of an attempt to undermine the dualism of first and second nature is uncharacteristic of his therapeutic approach. Though he acknowledges the major role of second nature without which we had no rational touch with the world, he does not manage to propose a way to undermine the misleading dualism. But that should be his goal, I claim: Since McDowell is right that the notion of first nature is flawed, he should want to propose an understanding that gets rid of it. I now want to use the remainder of this paper to develop this thought a bit further.

So how should an alternative look like? It helps us to take seriously the thoroughly therapeutic approach of McDowell's philosophy, which will show

how the dualism of first and second nature can be overcome. As I explained, McDowell himself does not make this last step, though it is still perfectly consistent with his philosophy.

The Wittgensteinian therapeutic approach that I follow McDowell in endorsing is best characterized in the metaphor of climbing the ladder which is to be thrown away after reaching the top (TLP 2019: 6.54). For the problem we are concerned with, this means the following: Originally, we started at a point where modern science gave us *one* notion of nature, which was said to explain the whole world around us. McDowell criticized this picture as being mistaken, as it is limited to what he calls the *disenchanted* aspects of nature, which the natural sciences are capable of explaining with their methodologies. What remains unexplained, then, entails large portions of concepts like reasons, values, and meaning. In order to draw these out of the space of the seemingly mysterious, McDowell introduces *second nature* and what he calls the *re-enchantment of nature* (McDowell 1996: 74), which allows us to widen the scope of our understanding. This move is part of the climbing of the ladder in the Wittgensteinian sense, as it cracks open a common but misleading understanding of nature and offers a more elaborate account.

However, the result of the McDowellian critique of first nature is not a point where we can already discard the ladder. Instead we are now confronted with two concepts of nature which partly coincide and overlap, but also compete against each other in some aspects: His continuous stressing of first nature as almost equivalent to the findings of the natural sciences not only maintains the gap between such and other explanations of the world, but also causes limited (if not inaccurate) conceptions of all that escapes such natural-scientific elucidation. Along with his dualistic picturing of the spaces of reasons and nature, this implies some hierarchy which gets reflected in his stressing of the need for a proper grip on the world, which in McDowell's argument seems to be maintained only as long as first nature shines through (Feige 2022: 167-170). So first nature—however much it is permeated by second nature—seems to remain a necessary anchor providing the starting point for our making sense of the world, and that still leaves us confronted with the problem of proper mediation.

This means that we are not yet at a point where things have calmed down, the Wittgensteinian therapy is not complete. There are still these two conceptions of nature which we need to set up so that they converge. To do so, we need to climb higher on the ladder to a point where we realize that we can finally throw it away; in order to reach again just *one* concept of nature. I am hesitant to call this a *unified* concept of nature since this is not exactly what it takes. We need no more work trying to bring first and second nature together in the sense that both conceptions merge. Instead, nature must be understood as something that somehow *entails* the aspects that first and second nature shed light on, but the way nature is talked about is not necessarily identical with either of the two present modes.

Let me illustrate the same point with a different metaphor: With Gadamer's hermeneutic circle in mind, we may view the original notion of (first) nature as providing the pre-conceptual starting point of our inquiry: At some point, then, we realize that first nature cannot explain all of what we thought it could; in Gadamer's words, we make "the experience of being pulled up short" (Gadamer 1990: 272; Gadamer 2004: 270; see also Lauer 2024: 171f.). McDowell's discussion of second nature is then, broadly speaking, part of the process of seeking improvement by going through the circle, which shall eventually provoke a revised understanding of the concept of nature. However, here again we can see that the investigation ends too early: As long as there are still two competing conceptions of nature present, we cannot speak of having made a full lap within the hermeneutic circle. We are stuck somewhere halfway through the circle, as we have only managed to shake up pre-existing conceptions by adding second nature to the discussion. At best we found a way to mediate the two, but only to such a limited extent that we could not yet reach a satisfactory state of a transformed understanding of nature. Once again such an understanding of nature could only be achieved once the circle is completed, when the process of hermeneutic understanding reaches a stage of temporary calm. McDowell has not yet gone that far since he keeps both conceptions of nature still in place. But only after reaching a conception of nature of the type I endorse we can talk about reaching a point of temporal calm. This is what should be our goal, and I assert that this is perfectly consistent with McDowell's thinking.

4. Conclusion

McDowell would certainly disagree with me if my point were merely to abolish the talk of first and second nature entirely, since such a conception gets close to endorsing bald naturalism, which aims at an *inclusion* of what we call the space of reasons into first-natural terminology (McDowell 1996: XXII), an obviously flawed characterization of what nature is. Wittgenstein demonstrates that we need a different way of untying those troublesome knots depriving us of an unobstructed view on nature. My conception is different from bald naturalism—and therefore consistent with McDowell's philosophy—as it elevates the term *nature* onto a stage of transformation: Acknowledging that nature *entails* all these aspects I discussed does not prohibit picking out different angles and approaches to talk about certain aspects of nature. This way, one could still talk about first and second nature to make distinctions if people find that necessary, but they did so within a practice that clearly marks these as merely provisional emphases that remain under the governance of nature as a whole.

A few issues may yet to be addressed elsewhere. The major challenge affiliated with my discussion so far is obvious: It is currently almost impossible to give any *positive* characterization of how my conception of nature can look like. We are stuck in some situation similar to what Heidegger states as causing his exclusively negative characterization of *being-in* (*In- Sein*), which he defends as follows:

The phenomenological demonstration of being-in-the-world has the character of rejecting distortions and obfuscations because this phenomenon is always already “seen” in every Da-sein in a certain way. And that is true because it makes up a fundamental constitution of Da-sein, in that it is always already disclosed, along with its being, for the understanding of being in Da-sein. (Heidegger 1996: 54; Heidegger 2006: 58)

The distortions that make it impossible to give a positive characterization of my conception of nature are a result of the long tradition of hermeneutic understanding involving deep entrenchments of first-natural scientific talking—an *always already seeing in a certain way*. It was possible to set the vocabulary of second nature against it because of a more narrow focus on the

questions that science leaves open and efforts to remember the concepts brought up long time ago. The result was the explicit conceptual distinction of the two natures, which is surely helpful, but caused the dualism we have been struggling with to come into being. But now that I propose to revise our understanding of nature by undermining the dualism, all we could characterize nature with is already part of the long tradition of either of the two understandings, and therefore all attempts to give a positive characterization of nature runs the risk of being misunderstood in just this way. We must make efforts to modify our *always already seeing in a certain way* to make it possible to understand nature in exactly the way I endorse.

My suggestion obviously has implications for several neighboring discussions as well. For example, a transformed conception of nature calls for a re-evaluation of the relation between nature and culture—in a very productive direction I suppose, since it may help us overcome that dualism once and for all. Understanding nature this way is also potentially relevant for reflections on the status of rationality in human infants and children, especially in opposition to animal nature. But there is no space to discuss these matters in this paper.

We see that what is needed is a committed engagement in the tradition of nature with a serious and continuous effort to re-connote these concepts we use on a daily basis. We need to make this effort in order to make partial steps towards reaching a transformed conception of nature, which can ultimately change the way we view the connection between the mind and the world. Wittgenstein and McDowell have accompanied us halfway through—now we must keep climbing the ladder until we really reach the top.

References

Bertram, Georg (2014) "Zweite Natur - Die Auflösung des Dualismus von Kultur und Natur", in: Christian Barth and David Lauer (eds.) *Die Philosophie John McDowells*, Münster: Mentis.

De Gaynesford, Maximilian (2004) *John McDowell*, Cambridge: Polity.

Feige, Daniel Martin (2022) *Die Natur des Menschen – Eine dialektische Anthropologie*, Berlin: Suhrkamp.

Gadamer, Hans-Georg (1990) *Wahrheit und Methode*, 6th edition, Tübingen: Mohr Siebeck.

Gadamer, Hans-Georg (2004) *Truth and Method*, 2nd, revised edition, Translated by Joel Weinsheimer and Donald G. Marshall, London, New York: Continuum.

Haddock, Adrian (2009) "McDowell, Transcendental Philosophy, and Naturalism", *Philosophical Topics* 37(1), 63-75.

Heidegger, Martin (1996) *Being and Time*, Albany: State University of New York Press.

Heidegger (2006) *Sein und Zeit*, 19th edition, Tübingen: Max Niemeyer.

Lauer, David (2024) "Coming Full Circle: Experience, Tradition, and Critique in Gadamer and McDowell", in: Daniel Martin Feige and Thomas J. Spiegel (eds.) *McDowell and the Hermeneutic Tradition*, New York: Routledge, 166-191.

McDowell, John (1996) *Mind and World*, Cambridge, Mass.: Harvard University Press.

McDowell, John (1998) "Two Sorts of Naturalism" in: John McDowell (ed.) *Mind, Value, and Reality*, Cambridge, Mass.: Harvard University Press, 167-197.

Müller, Jan (2017a) "Kritische Theorie und Aristotelischer Naturalismus", in: Martin Hähnel (ed.) *Aristotelischer Naturalismus*, Stuttgart: J.B. Metzler, 296-313.

Müller, Jan (2017b) "'Zweite Natur' und 'moderne Lebensform'", in: Jens Kertscher and Jan Müller (eds.) *Praxis und 'zweite Natur'. Begründungsfiguren normativer Wirklichkeit in der Diskussion*. Münster: Mentis, 151-164.

Rapp, Christof (2014) "McDowells Aristoteles", in: Christian Barth and David Lauer (eds.) *Die Philosophie John McDowells*, Münster: Mentis.

Testa, Italo (2007) "Criticism from within nature: The dialectic between first and second nature from McDowell to Adorno", *Philosophy & Social Criticism*, 33(4), 473-497.

Wittgenstein, Ludwig: *Werkausgabe Band 1. Tractatus logico-philosophicus / Tagebücher 1914–1916 / Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp 1984. [Neuaufgabe 2019]

Acknowledgement

I want to express my gratitude to David Lauer, without whom this paper wouldn't have been written. Thank you for your continuous support.

Is There a Plausible Realist Theory of Fictional Characters?

Andrew Wynn Owen (Austin, USA)

Abstract

The debate between realists and anti-realists about fictional entities is important partly because it connects with debates about the nature of reference. According to the descriptivist model held by Fregeans, a name has reference to an object due to the connection of that name with a description, which is met by the relevant object. According to the causal-communicative model held by Millians, a name refers in virtue of a chain of reference linking that name to a referent. In the case of fictional entities, it is a matter of debate whether the entities in question exist or not. Traditionally, Millians have had trouble dealing with anti-realism about fictional entities. I argue for a simple realist theory, 'bare-bones artifactualism', according to which fictional entities are simple abstract 'counters'. Each of these simple abstract counters has only those intrinsic properties that other such simple abstract counters also have, except for one distinctive intrinsic property. This one distinctive intrinsic property is a number-bearing property that marks out the abstract counter's identity, distinct from all other such abstract counters. The number-bearing property allows us to do with abstract counters what spatiotemporal situatedness allows us to do with concrete counters: it allows us to treat the abstract counters as individuals. In support of such artifactualism, I discuss the following considerations: its ontological simplicity; its parallels in our other practices; and its explanatory promise.

0. Introduction

The debate between realists and anti-realists about fictional characters is important partly because it connects with debates about the nature of reference. According to the descriptivist model of reference held by Fregeans (see Frege (1892) and Russell (1911)), a name has reference to an object due to the connection of that name with a description, which is met by the relevant object. According to the causal-communicative model held by Millians (see Mill (1867)), a name refers in virtue of a chain of reference linking that name to a referent. Kripke's (1972) work influentially revived the Millian view and challenged the Fregean view. In the case of fictional characters, it is a matter of debate whether the entities in question exist or not. If they exist (as possibilities that we discover, or as actual abstract artifacts that we create), the Millian view is on sturdy ground: there are referents to which the chain of reference leads. If they do not exist, more will need to be said by the Millian in order to account for what is going on in the case of apparent reference to such entities. Traditionally, Millians have had trouble dealing with anti-realism about fictional characters.

Here I will assess the plausibility of one realist view: a simple, bare-bones version of abstract artifact theory, which I call ‘bare-bones artifactualism’.

1. Bare-bones artifactualism

According to bare-bones artifactualism, the abstract artifacts are simple abstract ‘counters’. Each of these simple abstract counters has only those intrinsic properties that other such simple abstract counters also have, except for one distinctive intrinsic property. This one distinctive intrinsic property is a number-bearing property that marks out the abstract counter’s identity, distinct from all other such abstract counters. On account of this one distinctive intrinsic property, the abstract counters might be described as ‘quasi-counters’, given that, unlike idealized concrete counters, they are not indiscernible in their intrinsic properties. However, I will not describe them as ‘quasi-counters’, but rather ‘counters’, because it seems to me that their closeness to concrete counters is sufficient to justify categorizing them under that term. The one distinctive intrinsic property, the number-bearing property, allows for us to do with abstract counters what spatiotemporal situatedness allows for us to do with concrete counters. That is to say, the number-bearing property allows us to treat the abstract counters as individuals.

It is commonly thought that abstract objects cannot be distinct duplicates, unlike concrete objects. I accept that this is probably the case. So it is not possible for abstract objects to be perfect counters on the model of concrete counters, when ‘perfect counters’ is taken to mean ‘objects indistinguishable in their intrinsic properties’.

So the simple abstract counters are nearly indiscernible insofar as their intrinsic properties are concerned. Where the counters differ is in their extrinsic properties: these can be derived from an account of our use of them. By ‘counter’, I mean a simple object, indiscernible in its intrinsic properties from other such simple objects, that is used for tracking certain contextually-determined factors in a mutually-understood format (as in a game, or in our ordinary tracking of considerations about human society).

2. Is the ontology of bare-bones artifactualism strange?

Against the view that the ontology of bare-bones artifactualism using abstract counters is strange, I will present three considerations that give us good reason to believe in such an ontology: its simplicity; its parallels in our other practices; its explanatory promise. I will address each of these in turn, giving greatest attention to parallels in our other practices and explanatory promise.

2.1. Bare-bones artifactualism is ontologically simple

This is what I call ‘the simplicity advantage’. The proposed ontology is far simpler, and therefore more plausible, than the ontology of other versions of artifactualism: there are existent abstract counters, but their intrinsic properties are very simple and they are uniform in this simplicity. They are not spooky, because they are so simple in their intrinsic properties. Our ontology is minimally and plausibly expanded by acceptance of their existence, in keeping with the principle of parsimony.

2.2. Counter-use is familiar from other areas of human life

Bare-bones artifactualism using counters is plausible because there are analogies for such counter-use in other areas of human activity: notably, in games. When humans involve themselves in games, which may be very complex, they tend to have counters that they use to individuate possessors for the various properties in operation in those games. Moreover, in games, there is typically some ‘boardspace’ or ‘playing area’ within which the counters interact: we place counters on a board or cards on a table. Such areas, which are of significance to the use of counters in the game, are temporally extended: the counters enter them at a certain point in time, and leave at another. They are also, more obviously, spatially extended, allowing for various relations within that space, settled by the rules of the game. For example, counters sometimes preclude co-occupation of a square on a board, as in backgammon or chess. At other times, counters can co-occupy a square on a board, as in certain versions of pachisi (in which counters from the same team can co-occupy).

It should not be thought that counter-use is confined to the concrete. Consider two chess-players playing chess without a board. The game still takes place with counters, with various extrinsic, encoded properties (e.g. the knight moves in an L-shape, the bishop diagonally), but it takes place abstractly. Such

chess is sometimes said to be played ‘in the head’ and by this it is meant that it is played abstractly. The two players take it in turns to say their moves: ‘c4’, ‘e6’, ‘Knight to f3’, ‘d5’, and so on. When one plays (or attempts to play) such chess, one is aware of abstract counters and an abstract boardspace within which they operate. ‘Blindfold chess’, as it is sometimes known, has a long history: Murray (1913: 817) records that Ruy Lopez played it in Spain in the sixteenth century, and it probably goes back much further.

The same game-abstraction can be achieved for simpler games. The same abstract use of counters and boardspace can be achieved, for example, by children playing *Noughts and Crosses*. All it requires is some familiarity with the game. It seems probable that a person could play the abstract version of *Noughts and Crosses* (though probably not chess, due to the relative complexity of the game), without ever having seen a concrete boardspace or set of counters.

The widespread and commonsensically-recognized occurrence of such concrete and abstract counter-use in human life is a reason to view the further occurrence of abstract counter-use for fictional characters as plausible. It may be that, in the history of human evolution and/or in the course of human maturation from childhood to adulthood, either concrete or abstract counter-use precedes the other. My guess is that concrete counter-use precedes abstract counter-use, both in the history of human evolution and in the course of human maturation, but this is an area for further consideration, and one that is likely to be heavily dependent on empirical investigation. In any case, concrete counter-use preceding abstract counter-use (or vice versa) would not, as far as I can see, diminish the plausibility of fictional characters being simple abstract counters.

2.3 Counter-use has explanatory promise

As said, counter-use is ontologically simple and familiar from other areas of human life. But what explanatory work does counter-use do? Counter-use provides a solution to a coordination problem: the problem of how we organize our considerations of and discussions about fictional entities. Positing counters goes a long way to explaining how we achieve, where fiction is concerned, the kind of multi-character, multi-property, multi-event tracking that we do.

Imagine the game of *Snakes and Ladders* (known as *Moksha Patam*, in the original, Indian version) without counters, but with the aspects of the game that are brought about by the properties intrinsically or extrinsically possessed by counters in the normal game still in operation. How would we coordinate such complicated activity, without individuals to which to affix properties? In normal *Snakes and Ladders*, a counter has various extrinsic properties, provided by the game: when a die is rolled, it moves forward the relevant number of squares; when it lands on a square with a snake or a ladder, it respectively descends or ascends the board; when it reaches the end of the board, it results in victory for the associated player.

A version of *Snakes and Ladders* without the property-bearing, including spatiotemporal situatedness, provided by individual counters is either impossible or beyond ordinary human capabilities. I am not sure which of these is the case, but will outline the two possibilities that I envisage. It may be impossible: it may be that, without individual counters to bear the relevant properties, it would not be possible for the imitated game to functionally resemble the original. Alternatively, it may simply be beyond ordinary human capabilities: it may be that, even if it is possible for some other kind of being, our own capabilities are outmatched by the prospect of playing a version of the game that does not include individual counters to bear the relevant properties. On balance, I think it is probably impossible for there to be a version of *Snakes and Ladders* functionally identical to *Snakes and Ladders* and lacking the counters, but I am unsure on this point. In any case, certainty about this is not needed for my point about the implausibility of such a version of the game to go through.

Now imagine a novel in which there are no fictional characters, no counters to keep track of. I do not mean a novel in which various bits of speech and/or description are available, and the reader is challenged to gather them into world-resembling arrangements, and then does so by inventing individuals to which to attribute the various grouped patterns of speech and description. I mean something far more challenging: I mean a novel in which, at the outset, the reader has her capacity to track individual entities frozen. This is even stranger than the imitated version of *Snakes and Ladders*. It is hard to see how such a novel could be produced, let alone allow for our varied achievements in thought and discussion about fictional entities.

3. Objection: Is there an explanatory drawback concerning authorial creativity, due to the simplicity of the counters?

All versions of artifactualism have what I call ‘the creativity advantage’, which anti-realist and Meinongian realist accounts do not. On artifactualist views, the creator of a fictional character can genuinely be credited with having created something. Bare-bones artifactualism retains this advantage. However, there is a caveat. Such simple artifactualism may be seen as having this advantage to a lesser degree than versions of artifactualism that situate the created artifact as a more complicated object, one with more intrinsic properties. Simple abstract counters may not, on the face of it, appear to be the kind of thing that authors create: it may seem that what authors actually create are more complex, intrinsically-characterful abstract objects.

This is not a major drawback for bare-bones artifactualism because the apparently-missing explanatory work can be done by extrinsic properties. Imagine someone saying, as a challenge to the bare-bones artifactualist, ‘Simple abstract counters do not sound at all like what I think of fictional characters as being.’ Let us call this the ‘abstract-counters-are-the-wrong-things objection’. In answer to this objection, the bare-bones artifactualist can simply reply: ‘I don’t think that anyone, pre-reflectively, has ordinary or special insight into the ontological status of fictional characters.’

In fact, the bare-bones artifactualist is here using a defence that many artifactualists will probably adopt at some point, in response to a slightly different objection. That objection is raised by Sainsbury, as a concern about abstractness for artifactualism in general: ‘on abstract artifact theories, fictional characters just are not the kinds of things we want them to be’ because we do not think of fictional entities as abstract (2009: 111). I will call this the ‘abstracta-are-the-wrong-kinds-of-things objection’. In elaborating this objection, Sainsbury writes, ‘Authors, who ought to know, would fiercely resist the suggestion that they [fictional characters] are abstract. Abstract artifact theory entails that producers and consumers of fiction are sunk in error’ (111). Let us call this supporting point the ‘error hypothesis’. Sainsbury makes a normative claim about authorial knowledge (‘ought to know’) but the error hypothesis might also be couched as a simple statement of what is probable: it

seems probable that authors are not mistaken about fictional characters in this way. So there are normative and probabilistic versions of the error hypothesis. I will focus on the authorial, rather than the readerly, aspect of the error hypothesis, because it seems to me that, when the hypothesis is raised, it is to authors, rather than readers, that greater knowledge about what is going on with fictional entities is likely to be attributed.

In answer to the error hypothesis, I say that I do not think that authors ought to know what they, the authors, are doing in this regard; nor is it probable that they know. There is no special normativity and/or likelihood, for authors, when it comes to knowing the reality-status of fictional characters. In general, there are certain kinds of knowledge about the nature of the raw materials one is using in a project that are not required for successful completion of that project. For example, there is no special normativity and/or likelihood for those working on many aspects of the construction of buildings to know, simply by dint of engaging in those activities, the chemical composition of the materials that they are using. It may be that modern workers in that field do come to possess such knowledge, either as a matter of interest or in support of specialised aspects of their work, but it is not relevant to large aspects of the fundamental activity. For most of human history, the vast majority of those working on the construction of buildings did not have such knowledge, and yet the construction of the buildings was achieved. I believe there is an analogy here with what authors do. In fact, I think the analogy is generalizable to other domains: those working with certain raw materials, where the work involves the arrangement of those raw materials in time and space, and in other ways relative to one another (as in, for example, the social relations of a novel), typically are not likelier than the average person to know about the intrinsic nature of those raw materials. In general, I do not think that authors, mathematicians, or moralists have special insight into the debate about realism, anti-realism, and irrealism concerning the things with which they operate (be they fictional characters, numbers, or moral properties). These are distinctively philosophical questions and they do not routinely arise in pursuit of the relevant practices, though they are surely of interest for those concerned with foundational issues related to those practices.

4. Conclusion

I defended a view that I call ‘bare-bones artifactualism’, according to which the abstract artifacts created by authors are very simple: they are counters, which are indiscernible insofar as their intrinsic properties are concerned. Where they differ is in their extrinsic properties.

I said that I do not think the ontology of bare-bones artifactualism using abstract counters strange. On the contrary, there is good reason to see it as what is going on. This is based on certain considerations about bare-bones artifactualism: its ontological simplicity; its parallels in our other practices; and its explanatory promise. Above all, I would like to emphasize the account’s explanatory promise: abstract counter-use shows how we might be keeping track of the many things we have to be keeping track of in order to achieve the kinds of thought about and discussion of fictional entities that we do achieve.

I then addressed an objection: is there an explanatory drawback for bare-bones artifactualism, relative to other artifactualisms, due to the simplicity of the abstract counters? I said that the view that bare-bones artifactualism is at a disadvantage here relative to more elaborate versions of artifactualism is based on the incorrect supposition that there are categories of person who have special insight, where the debate about realism arises, into the ontological status of the objects of our thought and language.

It may be that the bare-bones model of artifactualism that I have outlined here remains unpersuasive. There is more that might be said about counter-use, but even an extended account might fail to convince the reader. For those who find even the existence of simple abstract counters implausible and yet also wish to retain the advantage (which is usually the preserve of the realist) of maintaining reference as something that we do in cases concerning fictional entities, Sainsbury’s (2009) irrealism is a good alternative view.

References

Capellen, Herman and Josh Dever (2018) *Puzzles of Reference*, Oxford University Press.

Frege, Gottlob (1892) "On Sense and Reference", in: P. Geach and M. Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell (1952).

Friend, Stacie (2007) "Fictional characters", *Philosophy Compass* 2 (2), 141–156.

Kripke, Saul (2013) *Reference and Existence: The John Locke Lectures*. Oxford, Oxford University Press.

Lewis, David (1983) "Extrinsic Properties", *Philosophical Studies*, 44, 197–200.

Mill, John Stuart (1867) *A System of Logic*, London: Longmans.

Murray, Harold James Ruthven (1913) *A History of Chess*, Clarendon Press, Oxford.

Russell, Bertrand (1911) "Knowledge by Acquaintance and Knowledge by Description", *Proceedings of the Aristotelian Society*, 11: 108–128.

Sainsbury, Mark (2009) *Fiction and Fictionalism* (1st ed.). Routledge. <https://doi-org.ezproxy.lib.utexas.edu/10.4324/9780203872567>

Thomasson, Amie (1999) *Fiction and Metaphysics*. Cambridge, U.K.: Cambridge University Press.

Van Inwagen, Peter (1977) "Creatures of Fiction", *American Philosophical Quarterly*, 14(4): 299–308.

Yagisawa, Takashi (2001) "Against Creationism in Fiction", *Philosophical Perspectives*, 15, 153–172.

Acknowledgement

Thanks to the University of Texas at Austin, and to Mark Sainsbury for discussion of this paper.

No Reduction of Slot-Filling to Grounding and Essence

Li Xiao (Vienna, Austria)

Abstract

Proponents of neo-Aristotelian mereology consider the concept of a “slot” pivotal for understanding the parthood relation. According to this view, an entity constitutes a part of a whole only in virtue of filling a mereological slot within that whole. Recently, Sattig (2021) has proposed that slot-filling can be defined through other metaphysical concepts, particularly grounding and essence. In this paper, I first distinguish between two conceptions of mereological slot: *parthood slot* and *hylomorphic slot*. Then I argue that his definition is over-restrictive with respect to *parthood slot*, which poses no restriction to the kinds of entities involved. And more importantly, his definition is over-general concerning both *parthood slot* and *hylomorphic slot*, in the sense of encompassing cases other than slot-filling. Additionally, I argue that the failure of this definition stems from a misguided direction of reduction: without sufficient conditions to distinguish parthood relationship, which is *one* way of grounding, from other ways, the attempt to reduce slot-filling into grounding is inherently over-general.

1. Introduction

The concept of “slot” has become popular among proponents of non-classical mereology. According to slot mereology, a composite whole does not simply possess parts in a straightforward manner, but instead, it has parthood slots filled by something. That is,

(Slot mereology) x is a part of y , at $t =_{\text{def}} y$ has a *slot* z and x *fills* z , at t .

The concept of a slot can be understood in both thin and thick senses. In the thin sense, a slot is merely a mereological position waiting to be filled, with no further requirement except parthood relations. Following Bennet (2013), I will call this kind of slots *parthood slots*. In contrast, the thick concept of slot entails specific requirements regarding the kind, feature, and/or relations of the potential filler. For instance, a car has a slot for an engine which can only be filled by a material object with particular features and connections to things of other kinds. Given its typical application in the context of neo-Aristotelian hylomorphism, which views material objects as composed of both form and matter, I will term those slots *hylomorphic slots*.

In a recent paper (Sattig 2021), Sattig endeavours to reduce the concept of *slot-filling* to essence and grounding. Suppose y is a whole with a slot that requires its filler to belong to a certain kind K (K -slot), and x fills this slot at time t ,

(Slot-filling) x fills a K -slot of y , at t , =_{def}

- (i) x is a K , at some region p , at t ;
- (ii) there is a kind K^* , such that y is a K^* , at some region p^* , at t ;
- (iii) x is *essentially* a K , while y is *essentially* a K^* ;
- (iv) the fact that some object is a K , at p , at t , *partially grounds* the fact that some object is a K^* , at p^* , at t ;
- (v) p is a *subregion* of p^* .

The rough idea of the definition lies in the notion that an object x fills a K -slot of y iff x is essentially a K , y is essentially a K^* , and the existence of a K at a space and a time partially grounds the existence of a K^* at an extended space and the same time. There is also a relational slot definition, but since it is based on the definition of K -slot, I will focus on the kind-based version.

While it may seem like progress to reduce the concept of slot-filling, typically considered primitive, to other more general metaphysical concepts like essence and grounding, I will demonstrate in this paper that Sattig's definition falls short. With regard to parthood slots, the kind restriction is not necessary and the left restrictions combined together are insufficient (Section 2). Concerning hylomorphic slots, the definition is overly inclusive, encompassing cases that we would not consider instances of slot-filling (Section 3). I will further argue that the attempt to reduce slot-filling to grounding and essence is misguided (Section 4). In sum, I will show that it is challenging to reduce the concept of slot/slot-filling to grounding/essence, and for the theoretical framework of both slot mereology and neo-Aristotelian hylomorphism, maintaining slot as a primitive concept is justified.

2. The problem with parthood slots

To start with, I will introduce the concept of a parthood slot according to Bennet (2013). Consider the familiar distinction between *roles* and *occupants*. For instance, Joe Biden occupies the role of the President of the United States. This role, established by the Constitution of the US, has been occupied by

individuals other than Biden in the past. Moreover, Biden occupies other roles as well. Different types of roles are characterized differently; for example, the role of the Presidency is defined by its responsibilities and powers, and the role of pain is characterized causally. Bennet proposes understanding the parthood relation in a similar way. A whole has several slots awaiting fulfillment, akin to roles characterized mereologically. That is,

To be a parthood slot is to be a location in a mereological nexus, to be an aspect of the mereological structure of a whole, to be defined by what it is part of and what is part of it. To occupy a parthood slot is to actually stand in the relevant parthood relations, to actually be part of some things and have other things as parts. (Bennet 2013: 87)

An object becomes a part of a whole in virtue of filling a parthood slot of that whole. Formally, x is a part of y iff y has a slot z and x fills z . We define the concept of parthood through *slot* and *slot-filling*.

Now it becomes apparent that Sattig's definition does not fit Bennet's framework. According to Sattig, for x to fill a slot of y , x must belong to a kind K . However, slot mereology in principle does not concern itself with the kind of filler occupying a slot. Bennet does not intend to confine the application of slot mereology to material objects with specific requirements to their parts. Rather, it is conceived as a general framework applicable to various types of parthood relations. As Bennet notes, even "classical extensional mereology can in fact be reformulated in these terms." (Bennet 2013, 87) Classical mereology imposes no restriction on the kind of parts. Therefore, Sattig's definition overly restricts parthood slot.

Certainly, Sattig could adopt a more flexible stance regarding the kind restriction. The kind of a part can trivially be *a thing* or *an object*, rendering the conditions (i)-(iii) effectively empty. However, matters become more complicated. To see this, we must scrutinize condition (v) more closely. Recall that x fills a K -slot of y only if x occupies the region p , y occupies the region p^* , and p is a subregion of p^* . According to Sattig, the region p is a subregion of the region p^* iff p is a *part* of p^* . The parthood relation in the definition is spatial parthood, and cannot be reduced to slot terminology (Sattig 2021, n. 8). The intuition behind this condition is as follows:

[T]he exact locations of a material object's parts, at a time t , are subregions of the exact location of the whole, at t . For instance, the engine, the chassis, the wheels, and all the other proper parts of a car, at a time t , are located inside of the region at which the car is located, at t . (Sattig 2021: 2739)

A wheel of a car occupies a region that is a part, indeed, *a proper part*, of the region occupied by the car. Then, why not define the subregion of a region as its *proper part* rather than just a *part*? I believe the reason lies in the following: Slot mereology accommodates cases where a whole has a part more than once. Suppose y has exactly two slots, a and b , and x fills both of them. In this case, x is a proper part of y . However, x occupies exactly the same region p as y does. According to condition (v), x occupies a subregion of the region that y occupies. As p is merely the improper part of p , we must define a subregion as a part, rather than a proper part, of a region.

Recall that to define parthood slot, we must trivialize conditions (i)-(iii). There are no kind restrictions on both part and whole. The weakened definition is as follows:

(Parthood-slot-filling) x fills a parthood slot of y , at t , =_{def}

- (i)* x exists at some region p , at t ;
- (ii)* y exists at some region p^* , at t ;
- (iv)* the fact that some object exists at p , at t , partially grounds the fact that some object exists at p^* , at t ;
- (v)* p is a subregion of p^* .

However, this definition faces counterexamples known as “interpenetration” (see e.g. Gilmore 2018, §4). Some metaphysicians believe in tropes, particularized properties/relations located at spatiotemporal regions. For example, suppose there is a red, round patch o at region p (which is a surface) at time t . According to trope theorists, there are tropes *the redness of o* and *the roundness of o* both precisely located at p at t . Moreover, it is not implausible that an object has a color partially in virtue of its having an extended shape. And o has an extended shape in virtue of being round. Thus, the fact that *the roundness of o* exists at p at t partially grounds the fact that

tropes *the redness of o* exists at *p* at *t*. The region of *p* is a subregion of itself. Hence, the two tropes satisfy all the conditions mentioned above. However, we do not say that *the roundness of o* is a part of *the redness of o*. The two tropes do not overlap. Consequently, the weakened definition of slot-filling also fails, this time being over general.

The upshot of this section is that Sattig's definition does not align with Bennet (2013)'s conception of parthood slots. But Sattig clearly mentions Bennet's paper (Sattig 2021, n. 4). The issue arises from the fact that the concept of parthood slot is meant to be general, devoid of requirements concerning the kind of whole and part. In this sense, Sattig' definition is overly narrow. Yet, once removing the kind requirements, the resulting definition becomes overly broad, encompassing pairs between which no parthood relations hold. In the next section, I will demonstrate that the problem of over-generality reappear concerning hylomorphic slots.

3. The problem with hylomorphic slots

While Sattig's definition of slot-filling fails to capture the concept of parthood slot, one might expect it to effectively address the thick concept of slot, i.e. hylomorphic slot. According to neo-Aristotelian mereology, a whole of a kind K^* has a K -slot that requires its filler to belong to kind K . For example, a car has an engine-slot which can only be filled by an engine, a H_2O molecule has an oxygen-slot reserved for an oxygen atom. Conditions (i)-(iii) of Sattig's definition now appear to be necessary. However, I will show that the problem of over-generality resurfaces.

Consider the following counterexample. Let (i) x be a brain event of kind N (e.g. a brain neural firing) occurring at p at time t , and (ii) y be a mental event of kind M (e.g. feeling pain) occurring at p^* at time t . I assume that (iii) x is essentially a N brain event and y is essentially M . Furthermore, suppose that the occurring of x , along with other brain events, realizes the mental event M for a human being at time t . Thus, (iv) the fact that x is N at t partially grounds the fact that y is M at t . Event x occurs at a region p that a portion of the brain occupies at t . Event y occurs at least at region p^* containing p at t . Consequently, (v) p is a subregion of p^* . Events x and y thus fulfill all five

conditions of Sattig's definition. However, does x fill a N -slot of y and qualify as a part of y ? I contend that the answer is negative. The relationship between x and y is not one of parthood but rather realization. The contrast becomes clearer when we consider the total brain event z that realize y at t . There is a sense in which x is a part of z , for x is one of the events that combine to form the total event z . However, z is not a part of y ; rather, z *realizes* y . Consequently, even though x is a part of z , it is not a part of y . Instead, x is *part of the realizer* of y .

Some may argue that the aforementioned case fails to satisfy the conditions, contending that condition (iv) stipulates "the fact that some *object* is a K , at p , at t , partially grounds the fact that some *object* is a K^* , at p^* , at t^* ". But in the case described, x and y are *events*, not *objects*. In response, I argue as follows: First, the concept of an object is vague and can encompass a wide range of entities. While in a narrow sense, e.g. material objects, events may not be considered objects, in a wider sense, including whatever can be the value of first-order variables (see Davidson 1967), events can be considered objects. Second, even if Sattig has a narrower sense of object in mind, the case illustrates the limitation of his definition. As previously observed, it makes sense to discuss the parts of an event. Furthermore, in discussions of neo-Aristotelian mereology, some philosophers explore the part of events in addition to the part of material objects (e.g. Fine 1982). Given the parallel between parthood of material objects and events, this case serves as a cautionary indication of a potential flaw in Sattig's definition.

Before delving into general discussions, I will present a potential counterexample involving material objects. Some argue that, according to physics, two bosons can occupy exactly the same spatiotemporal point without overlapping (Hawthorne and Uzquiano 2011, 55-6). Additionally, there exist entangled particle pairs, where one particle is in a spin-up state iff the other is in a spin-down state (Schaffer 2010, 51). Now, it is *conceivable* that a pair of co-located and entangled particles exists, where one is in a spin-up state and the other is in a spin-down state. For the up particle, its being up is at least partially grounded in the down particle's being down, and vice versa. But neither the up particle nor the down one is a part of the other.

One concern regarding this case is that neither particles is essentially up or down *de re*. But condition (iii) mandates that both entities essentially belong to their respective kinds. Nevertheless, let us consider a typical material object, such as a statue *y*, which has a hand-shaped part *x*. According to Sattig, *y* is essentially a statue and *x* is essentially hand-shaped. However, is *y* essentially a statue *de re*? If we understand *y* as merely the material substance, the answer is negative. Instead, we should understand Sattig as positing that *y qua statue* is essentially a statue. For something to be a statue, it must possess a hand-shaped part. Consequently, a thing has the hand-shaped slot only when considered as a statue; otherwise, it would not require one of its part to be hand-shaped. Similarly, for something to be an up particle in an entangled pair, it must have a down counterpart. The up particle in the entangled pair *qua up particle* is essentially up. When understood in this manner, the case serves as a counterexample to Sattig's definition.

4. The diagnosis

In this concluding section, I will briefly address what I perceive to be the fundamental flaw in Sattig's definition. In a nutshell, the direction of reduction is misguided. His underlying notion is to define that *x* fills a mereological slot of *y* as that *x*'s (essentially) being *K* partially grounds *y*'s (essentially) being *K**, while *x* is located in a subregion of *y*'s. However, grounding is a general metaphysical connection encompassing various relationships like parthood, realization, determinate-determiner, etc. Hence, it is entirely possible for a fact involving *x* to ground another fact involving *y*, without the grounding occurring via the parthood relation. For instance, in the context of neural and mental events, the way of grounding is realization. Similarly, in the case of tropes and entangled particles, the grounding mechanisms diverge. Consequently, the general concept of grounding proves too coarse-grained to capture the more nuanced concept of parthood. The reduction can only succeed with additional conditions sufficiently distinguish parthood from other ways of grounding. Unfortunately, the condition of subregion cannot address all counterexamples because the relations of parthood and spatial location do not necessarily align in the usual manner.

I would like to differentiate my skepticism about reducing slot-filling to grounding from Wilson's (2014) general skepticism regarding the utility of the concept of grounding. My contention is that grounding is too coarse-grained to encapsulate the specific parthood relation of interest in slot mereology. However, I don't reject the idea that we can grasp the features of grounding in a general sense, and these general insights can offer some illumination in our specific metaphysical inquiries.

5. Conclusion

In this paper, I argue that Sattig's attempt to define slot-filling using the concepts of grounding and essence falls short. Initially, I demonstrated that the kind restriction proves too narrow to capture parthood slots adequately. Furthermore, the issue of over-generality arises both in the weakened definition of parthood slot and the original definition concerning hylomorphic slots. My diagnosis is that the crux of the problem lies in the misguided direction of his reduction, aiming to reduce a fine-grained concept into coarser-grained ones.

References

- Bennett, Karen (2013) "Having a part twice over", *Australasian Journal of Philosophy* 91, 83–103.
- Davidson, Donald (1967) "The Logical Form of Action Sentences", in: Nicholas Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh (PA): University of Pittsburgh Press, 81–95.
- Fine, Kit (1982) "Acts, Events and Things" In Werner Leinfellner, Eric Kraemer and Jeffrey Schank (eds.) *Language and Ontology: Proceedings of the 6th International Wittgenstein Symposium*, Vienna: Hölder-Pichler-Tempsky, 97–105.
- Gilmore, Cody (2018) "Location and Mereology" In Edward N. Zalta (ed.) *The stanford encyclopedia of philosophy* (Fall 2018 Edition). <https://plato.stanford.edu/archives/fall2018/entries/location-mereology/>
- Hawthorne, John and Uzquiano, Gabriel (2011) "How Many Angels Can Dance on the Point of a Needle? Transcendental Theology Meets Modal Metaphysics", *Mind* 120, 53–81.
- Sattig, Thomas (2021) "Part, slot, ground: foundations for neo-Aristotelian mereology", *Synthese* 198 (Suppl 11), 2735–2749.
- Schaffer, Jonathan (2010) "Monism: The Priority of the Whole", *Philosophical Review* 119, 31–76.

Wilson, Jessica (2014) “No Work for a Theory of Grounding”, *Inquiry: An Interdisciplinary Journal of Philosophy* 57, 535-579.

Acknowledgement

I would like to thank Benjamin Schnieder, Damian Aleksiev and Yannic Kappes for helpful discussion.

Is Aboutness Commitment Possible?

Peihong Xie (Vienna, Austria)

Abstract

This paper introduces and evaluates a possible truthmaking argument for a robust conception of aboutness commitment. According to the argument, the pure aboutness-based account of truthmaking and the thesis of Truthmaker Commitment entails that a truthbearer is ontologically committed to what it is about in an appropriate sense of “aboutness”. However, this argument encounters challenges, notably an inherent explanatory circularity within the robust conception of aboutness commitment, and a radical inconsistency between its two premises in terms of their underlying motivations. To address these issues, this paper proposes a more moderate conception of aboutness commitment, drawing inspiration from B. Smith’s mixed aboutness-based account of truthmaking. According to this moderate proposal, the representationalist notion of intended aboutness, or the anti-representationalist notion of general aboutness, constrains, but does not determine, the notion of ontological commitment, which in turn contributes to a stronger notion of genuine or satisfied aboutness. This modified moderate conception not only sidesteps the explanatory circularity, but also aligns more closely with the objective of providing a fundamental ontology. Therefore, it is recommended as a fitting approach to engaging in the relation between aboutness and ontological commitment.

1. Introduction

This paper explores the prospect of aboutness commitment from the perspective of truthmaking, or of what portion of reality makes what truthbearers (namely representational entities with truth-values) true. What I mean by “aboutness commitment” is roughly the idea that a truthbearer is ontologically committed to what it is about. Put differently, what is required to be a portion of reality for a truthbearer to be true is provided by the target of aboutness of the truthbearer. At first glance, “aboutness commitment” in this sense is intuitively implausible considering lots of cases where a sentence is about, but not ontologically committed to, something. For example, it seems to be the case that “Pegasus is a mythological winged horse” is about Pegasus, which is ordinarily deemed as non-real (Crane 2013: 120). Or, suppose that “Axel” names a particular future adult human being who will be born after 300 years, it seems to be the case that “Axel is a human being” is about, but not ontologically committed to, Axel. Particular ontological debates over fictional or future entities do not matter here. For anything non-real, say, x , it is possible in principle to construct a truthbearer about, though clearly not ontologically committed to, x .

Nevertheless, the idea of aboutness commitment is not so easily refuted, especially when we realize that there are two independently plausible theses—the pure aboutness-based account of truthmaking and the thesis of truthmaker commitment—that could be used for motivating it. I will present a possible truthmaking argument for a robust conception of aboutness commitment, show why it is deeply problematic, and finally propose a more moderate conception of aboutness commitment as a remedy.

2. Motivating a Robust Conception of Aboutness Commitment

To see what an aboutness-based account of truthmaking is like, it is good to start with B. Smith's (1999) notion of *projection*. Smith argues against the *simple necessitation account* of truthmaking, the view that a portion of reality makes a truthbearer true if and only if the former's existence necessitates the former's truth. His counterexamples appeal to so-called "*malignant necessitators*". For example, the existence of Restall's refrigerator necessitates the truth of Goldbach's conjecture (if it is true), but Restall's refrigerator is not generally taken as a candidate truthmaker for Goldbach's conjecture (Smith 1999: 283). After all, this case of necessitation is "malignant" because Goldbach's conjecture, if true, is a necessary truth and thus trivially necessitated by anything irrelevant to its truth.

In light of this, Smith recommends to take *projection* as a relevance constraint on truthmaking. That is, a portion of reality serves as a truthmaker for a true truthbearer not only because the former necessitates the latter, but also because the former is projected by the latter. Smith intends his notion of projection as an explication of the general notion of aboutness. As he writes, "A truthmaker for a given judgment ... must be part of that which the judgment is *about*" (Smith 1999: 279). In a later paper, he also claims that his theory "captures the idea that p , if it is to be made true by x , must be about x " (Smith 2002: 231). Return to the above case of Goldbach's conjecture: The reason why Restall's refrigerator is not a truthmaker for Goldbach's conjecture is that the former is not projected by the latter, or that the latter is not about the former in an appropriate sense of "aboutness".

Formally speaking, Smith's notion of projection is characterized by *modal entailment*: x is projected by p iff p is true and p entails the existence of x . Put

aside whether this characterization of projection is appropriate, it immediately reminds us of the *entailment account* of ontological commitment, which is the general view that a truthbearer is ontologically committed to *Ks* iff it entails that *Ks* exist (Bricker 2016). This similarity is probably why B. Schnieder (2006) and A. Jiang (2023) both judge Smith's notion of projection as an explication of the notion of ontological commitment. If their judgment is right (I will argue against it later), the notion of aboutness would be closely connected to the notion of ontological commitment, revealing the conceptual possibility of aboutness commitment.

One or another version of the pure aboutness-account of truthmaking has been developed by truthmaking theorists such as P. Audi (2020) and A. Schipper (2018, 2020) since Smith's seminal work. Their accounts of truthmaking are *pure* because truthmaking are regarded as a *pure semantic phenomenon* in the sense that a portion of reality is a truthmaker for a true truthbearer iff the portion of reality is what the truthbearer is about in an appropriate sense of "aboutness". When combined with another promising thesis of *truthmaker commitment* defended by D. M. Armstrong (2004) and R. Cameron (2008, 2010), the pure aboutness-account of truthmaking will give us a plausible argument for aboutness commitment as follows:

The Truthmaking Argument for Aboutness Commitment

(1) What makes a truthbearer true if it is true is exactly what it is about (in an appropriate sense of "aboutness").

(2) *Truthmaker Commitment*: A truthbearer is ontologically committed to what makes it true if it is true.

Therefore,

(3) A truthbearer is ontologically committed to what it is about (in an appropriate sense of "aboutness").

In the above argument, the premises (1) and (2), though somewhat plausible, are still controversial, respectively. But what matters to us is whether their combination is promising in justifying the robust conception of aboutness

commitment, or at least providing some reason for showing that it is less implausible as it initially seems. My answer is “No,” based on my evaluation in the following two sections.

3. Representationalism versus Anti-representationalism

The plausibility of the premise (1) and thus the conclusion in the truthmaking argument rests on whether there is a well-justified connection from the aboutness of a truthbearer to a portion of reality (as the truthmaker for the truthbearer and thus as what the truthbearer is ontologically committed to). In this regard, *representationalists* in Schipper’s (2018: 3707) sense will deny that there is any truth *really* about anything non-real such as Pegasus on the grounds that aboutness in their view is always a genuine relation that involves real relata. For example, Smith and Ceusters (2015: 3) argue, “the aboutness in question must always be veridical, so that ‘being about’ is a success verb.” If Pegasus is non-real, we may have a fictional truth like “Pegasus is a mythological winged horse” that is *intended to be about* Pegasus, though its intention fails in that case.

By contrast, *anti-representationalists* like Crane (2013) and Schipper (2018) argue that aboutness is not always a genuine relation, thus allowing a truthbearer to be really about non-reality. Nevertheless, there is still some way available to them in establishing the connection from aboutness to reality. For example, Schipper (2018: 3704) introduces the notion of *aboutness-satisfaction*. The aboutness of a truthbearer is not satisfied if what it is about is not a portion of reality. In that case, the truthbearer is a *truthmaker gap* so that it lacks a corresponding truthmaker even if it is true. But if its aboutness is satisfied, then a genuine relation of aboutness would be instantiated, and thus the target of its aboutness would provide a truthmaker for its truth. So a revised form of the truthmaking argument for aboutness commitment would be available to anti-representationalists if the notion of aboutness both in the premise (1) and the conclusion (3) is replaced with the notion of *satisfied aboutness*.

In light of this, whether one is a representationalist or not, it seems that she can always somehow connect genuine or satisfied aboutness to reality. Then, are the premise (1) and the robust conception of aboutness commitment

thereby well-justified assuming the plausibility of the premise (2)? I doubt so. For aboutness commitment, if it has a good claim to being a conception of ontological commitment, is supposed to provide us with a test or method used for identifying what a truthbearer has to be ontologically committed to for it to be true. Put another way, no candidate conception of ontological commitment should be premised on any prior assumption of what is real in one's ontology. But for the representationalist notion of aboutness to work in identifying the ontological commitment of a truthbearer, we would have to first determine what is real, otherwise we would only have a case of intended aboutness but not of genuine aboutness. Similarly, if we identify the ontological commitment of a truthbearer using the anti-representationalist notion of satisfied aboutness, we would have to first determine when the aboutness of the truthbearer is satisfied; that is, we would have to figure out what there is in reality before implementing the supposed ontological test provided by aboutness commitment.

Therefore, the idea of aboutness commitment, whether interpreted from a representationalist or anti-representationalist viewpoint, fails to provide a useful test of ontological commitment due to its inherent explanatory circularity. This does not mean that there is anything incoherent in attempting to connect aboutness to reality. But it shows that any credible approach to understanding this connection has to presuppose some prior conception (and criterion) of ontological commitment.

4. Truthmaker Commitment and Fundamentality

To see another potential worry facing the robust conception of aboutness commitment, it is time to consider the premise (2) that a truthbearer is ontologically committed to what makes it true if it is true. This thesis is mostly proposed by defenders of the ontological approach to truthmaking, which is the view that a portion of reality plays its truthmaker role by virtue of its *ontological contribution* to the truth of the truthbearer it makes true. It is the supposed ontological contribution that explains the core insight of truthmaking that the truthbearer depends on the truthmaker, but not vice versa.

Although there are multiple ways of understanding what the ontological contribution of a truthmaker is (for example, J. Asay (2020), though somewhat unconvincingly, regards it as the necessitation of the truth of a truthbearer by the existence of its truthmaker), a promising proposal inspired by J. Schaffer (2008) and Cameron (2010) is to understand it as providing an *ontological ground* for the truth of the truthbearer in question. In this way, the thesis of Truthmaker Commitment is connected to the notion of *fundamentality* because ontological commitments, according to the proposal, must be provided at least by the truthmakers that ontologically grounds the truth of the corresponding truthbearers and thus are more fundamental than those truthbearers. In advocating for a “radically minimal ontology,” Cameron (2010: 250) even goes further to argue that what plays the truthmaker role can only be the entities with real being, or the entities at the ground level of the hierarchy of reality. So ontological commitments in Cameron’s sense are commitments to the absolutely fundamental entities.

However, no proponent of the pure aboutness-based account of truthmaking would agree with the above proposal, let alone its Cameronian version. As Audi (2020: 569) nicely presents, the basic idea behind the pure aboutness-based account of truthmaking is truthmaking as *matching*. That is, a truthmaker for a truthbearer plays its role not in terms of any ontological contribution it makes, but rather in terms of the semantic features of the truthbearer highlighted in the guise of aboutness. Audi says, “Truth is guaranteed once we have a truthbearer, the semantic features that constitute its matching a certain fact, and the obtaining of that fact” (Audi 2020: 583). In making a truthbearer true, all substantive work is done by its semantic character so that the obtaining fact is merely an “inert” participant (Audi 2020: 574-575). No further ontological contribution from the truthmaker fact is needed in truthmaking.

Since the proponents of Truthmaker Commitment and the theorists of pure aboutness-based truthmaking radically disagree on whether a truthmaker is supposed to make any ontological contribution in the enterprise of truthmaking, it is difficult to see how the premises (1) and (2) could be coherently used for establishing the conceptual possibility of aboutness commitment in the robust sense, even though the whole truthmaking argument is formally valid. On the one hand, given the proposal concerning

ontological ground, Truthmaker Commitment is intended to provide a fundamental ontology. On the other hand, however, the pure aboutness-account of truthmaking is not sensitive to the fundamental/non-fundamental distinction. No appropriate “aboutness apparatuses”—in Schipper’s (2018) locution—could be used for identifying what is relatively or absolutely fundamental, even if some of them might signify what there is.

Indeed, there is no denying that the proponents of the pure aboutness-account of truthmaking are free to reject the objective of fundamental ontology implied in Truthmaker Commitment, thus accommodating non-fundamental aboutness commitments. But this move may undermine the metaontological role of truthmaking in providing possible *ontological explanations* for truths. If Smith and Simon (2007: 90) are right when they say, “A theory ontologically commits us to those entities whose existence is required to *ontologically explain* its constituent assertions,” a concession to non-fundamental commitments would probably impose a theoretical burden on the robust conception of aboutness commitment itself, too.

5. Aboutness and Ontological Commitment

If the truthmaking argument for aboutness commitment fails in the abovementioned ways, one may ask, what is a correct understanding of the connection of aboutness and ontological commitment?

Let us reconsider the story of Smith’s (1999) notion of projection. Smith does think that an appropriate account of truthmaking has to be constrained by the element of aboutness for the purpose of precluding irrelevant truthmakers, but does not thereby adopt a pure aboutness-based account of truthmaking. Rather, his account of truthmaking is *mixed*: Truthmaking is regarded as a matter of necessitation of truthbearer by truthmaker plus aboutness. Indeed, there has been generally accepted among truthmaking theorists since G. Restall (1996) that there are well-known difficulties concerning the project of characterizing truthmaking in terms of necessitation of truthbearer by truthmaker. But Smith’s mixed account of aboutness is still insightful if the element of necessitation is merely seen as a (though inadequate) way of capturing the ontological contribution of a truthmaker. That is, aboutness is not the whole story of truthmaking. Instead, truthmaking is essentially an

enterprise of ontological explanation filtered through aboutness: Under the constraint of aboutness, a truthmaker for a truthbearer is regarded as ontologically explaining the truth of the truthbearer by providing a relevant and legitimate ground for it. Since the ontological commitment of a truthbearer is exactly what is needed by an appropriate ontological explanation of its possible truth, the truthbearer has to be ontologically committed to its truthmaker.

Such a general account of truthmaking is enough for motivating Truthmaker Commitment, while aboutness is still allowed to play a role in ontological commitment at the same time. Since a truth is ontologically committed to its truthmaker, and its truthmaker is subject to the constraint of aboutness, its ontological commitment would be constrained by aboutness so that what it is ontologically committed to must be within the scope of what it is about. To avoid the explanatory circularity mentioned in the section 3, the notion of aboutness can be weakened so that it is not required to have an indispensable connection to reality anymore. In other words, the moderate notion of aboutness involved in constraining truthmaker and ontological commitment is either the notion of intended aboutness for representationalists, or the notion of general aboutness (not necessarily requiring satisfaction) for anti-representationalists.

To further clarify the moderate conception of aboutness commitment, consider a simple argument from Jiang (2023) that that “is ontologically committed to” and “is about” are not co-extensive. Jiang motivates his point using the two following cases (Jiang 2023: 2):

Case 1: <Every swan can fly> is about swans, but not ontologically committed to them.

Case 2: <Anna is singing> is ontologically committed to human cells (assuming that Anna is a human being), but not about human cells.

It is not difficult to see that the Case 1 does not threaten the proposed moderate conception of aboutness commitment, because the aboutness constraint imposed on ontological commitment is allowed to be insufficient.

On the other hand, however, the Case 2 clearly violate our requirement that what a truthbearer is ontologically committed to must be within the scope of what it is about.

Nevertheless, the Case 2 is itself rather doubtful: I agree with Jiang that <Anna is singing> is not about human cells, but do not see why it may be ontologically committed to human cells by itself. A core reason provided for his claim is, “It cannot be true unless human cells exist” (Jiang 2023: 2). But such a reason is not well-grounded unless some particular background conditions concerning a human being and her relation to her cells are assumed. Even if <Anna is singing> requires Anna’s existence, it is another substantive ontological matter whether her existence requires the existence of her cells. This ontological matter cannot be adjudicated on the basis of the truth of <Anna is singing> alone. Therefore, Jiang does not successfully justify the separation of ontological commitment from aboutness.

6. Conclusion

As we have seen, a promising moderate conception of aboutness commitment is hidden behind Smith’s mixed aboutness-based account of truthmaking. That is, aboutness in the proposed moderate sense is a necessary but insufficient constraint for ontological commitment. The moderate notion of aboutness is responsible for constraining, but not determining, the notion of ontological commitment, which in turn contributes to our understanding of a stronger notion of aboutness in the genuine or satisfied sense. Such a picture of aboutness and ontological commitment is immune from the explanatory circularity facing the previous robust conception, while respecting both the objective of providing an (at least relatively) fundamental ontology implied in Truthmaking Commitment and the general insight behind a (mixed) aboutness-based account of truthmaking.

References

- Armstrong, David M (2004) *Truth and Truthmakers*, Cambridge: Cambridge University Press.
- Asay, Jamin (2020) *A Theory of Truthmaking: Metaphysics, Ontology, and Reality*, New York, NY: Cambridge University Press.

Audi, Paul (2020) "Why Truthmaking Is Not a Case of Grounding", *Philosophy and Phenomenological Research* 101(3), 567-590.

Bricker, Phillip (2016) "Ontological Commitment", in: Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), URL = <<https://plato.stanford.edu/archives/win2016/entries/ontological-commitment/>>.

Cameron, Ross (2008) "Truthmakers and Ontological Commitment: Or, How to Deal with Complex Objects and Mathematical Ontology without Getting into Trouble", *Philosophical Studies* 140, 1-18.

Cameron, Ross (2010) "How to Have a Radically Minimal Ontology", *Philosophical Studies* 151, 249-64.

Crane, Tim (2013) *The Objects of Thought*, Oxford: Oxford University Press.

Jiang, Asher (2023) "Truthmaking Cannot be Done afar", *Synthese* 201(3), 1-17.

Restall, Greg (1996) "Truthmakers, Entailment and Necessity", *Australasian Journal of Philosophy* 74(2), 331-340.

Schaffer, Jonathan (2008) "Truthmaker Commitments", *Philosophical Studies* 141(1), 7-19.

Schipper, Arthur (2018) "Aboutness and Negative Truths: A Modest Strategy for Truthmaker Theorists", *Synthese* 195(8), 3685-3722.

Schipper, Arthur (2020) "Aboutness and Ontology: A Modest Approach to Truthmakers", *Philosophical Studies* 177(2), 505-533.

Schnieder, Benjamin (2006) "Troubles with Truth-making: Necessitation and Projection", *Erkenntnis* 64(1), 61-74.

Smith, Barry (1999) "Truthmaker Realism", *Australasian Journal of Philosophy* 77(3), 274-291.

Smith, Barry (2002) "Truthmaker Realism: Response to Gregory", *Australasian Journal of Philosophy* 80 (2), 231-234.

Smith, Barry; Ceusters, Werner (2015). "Aboutness: Towards Foundations for the Information Artifact Ontology" In *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO)*, CEUR, Vol. 1515, 1-5.

Smith, Barry; Simon, Jonathan (2007) "Truthmaker Explanations", in: Jean-Maurice Monnoyer (ed.) *Metaphysics and Truthmakers*, Ontos Verlag, 79-98.

Contingentism, Grounding and the Reduction of Metaphysical Necessity to Essence

Qichen Yan (Princeton, USA)

Abstract

Trevor Teitel (2019) argues that the following three doctrines are jointly inconsistent: i) the doctrine that metaphysical necessity reduces to essence; ii) the doctrine that possibly something could fail to exist; and iii) the doctrine that metaphysical necessity obeys the modal logic of S5. This paper presents a novel solution to Teitel's puzzle, which is crucially based on the idea that the putative reductive relation between metaphysical necessity and essence should be understood as a type of *grounding* relation, rather than a type of *identity* relation. It will be shown that once we embrace this weaker version of the reduction (and some plausible assumptions about grounding), Teitel's puzzle can be naturally resolved. It will also be shown that the proposed ground-theoretic account has several advantages over its rival accounts.

1. Introduction

Following Fine (1994), many contemporary metaphysicians hold that for a proposition to be metaphysically necessarily true is for it to be true in virtue of the nature of some entities. However, in a recent paper, Trevor Teitel argues that so long as one endorses the following two principles about the reduction of essence to necessity, one must reject either contingentism (the thesis that possibly something could fail to exist) or the claim that metaphysical necessity obeys a modal logic of S4 (not to mention S5):

Canonical Reduction: $\Box\varphi$ reduces to $\exists xx \Box_{xx} \varphi$

Robustness: If the canonical reduction to essence is true, then it's necessarily true, necessarily necessarily true, and so on.

Here, "xx" denotes a *plural* free variable; " \Box_{xx} " denotes the essentialist operator "it lies in the nature of the plurality xx that". So the sentence " $\exists xx \Box_{xx} \varphi$ " can be read as "it lies in the nature of some objects that φ ".

It is noteworthy that **Canonical Reduction**, as it stands, stays neutral on what counts as a genuine *reduction* of $\Box\varphi$ to $\exists xx \Box_{xx} \varphi$. According to Teitel, this putative reductive relation can be understood in many different ways: it might be a type of *identity* relation; it might be a type of *grounding* relation; or it even might be a type of *linguistic/metasemantic* relation--in the sense of Sider(2011).

Teitel thinks that his argument against the combination of contingentism, the modal logic of S5, and the reduction of metaphysical necessity to essence has no bearing on which conception of reduction we appeal to. The goal of this paper is to show that Teitel is wrong: even if his argument is valid in cases where the putative reduction is underwritten by an *identity* relation, it is not a threat to essentialists who understand this reduction in terms of some *weaker* reductive relation, such as *grounding*. To show this, I will begin by introducing Teitel's argument in Section 2 and then examine in Section 3 a recent solution to Teitel's puzzle, which is given by Werner (2020). I will argue that Werner's solution fails, but his effort to address Teitel's puzzle is headed in the right direction. Then, in Section 4, I will present my own account, which is formulated through appealing to *grounding*. I will show how this ground-theoretic account can both resolve Teitel's puzzle and avoid the problems faced by Werner's account.

2. Teitel's Puzzle

Teitel's original presentation of the puzzle involves quantifications into sentence position. Using "p" as a propositional variable, he invites us to consider the following principle:

Standard Contingentism: $\exists p \diamond (\exists xx \Box_{xx} p \wedge \diamond \neg \exists xx \Box_{xx} p)$

Standard Contingentism says that some propositions are possibly both essential to some objects and possibly essential to nothing. This is an intuitive thought. The proposition that *if Socrates exists then he is a human* just is such an example: in a possible world where Socrates exists, this proposition is essential to Socrates, but in a possible world where Socrates does not exist, this proposition seems to be essential to nothing—provided that every world in which Socrates does not exist is a world in which no *higher-order* entity *about Socrates* could exist. Rejecting this latter condition amounts to embracing *hybrid contingentism*—the thesis that although it is contingent what individuals there are, it is not contingent what propositions and properties there are. This is because on a hybrid account, even if Socrates does not necessarily exist, his haecceity (i.e., the property of being Socrates) and his haecceity proposition (i.e., the proposition that Socrates is identical to himself) necessarily exist. So it is tempting to think that necessarily, there are some higher-order entities about Socrates (namely Socrates's haecceity and haecceity proposition) that

are essentially such that *if Socrates exists then he is a human*. Generalizing this idea will consequently give us a principled reason to reject **Standard Contingentism**.

Teitel concedes that his argument is not a problem for hybrid contingentists. So in this paper I will assume *higher-order contingentism*---the view that it is contingent both what individuals there are and what properties and propositions there are. Given this assumption, any contingentist who takes the putative reductive relation between metaphysical necessity and essence to be an *identity* relation will have some difficulty in accommodating the claim that metaphysical necessity's modal logic is at least S4. To see this, let's use " $\Box p$ " to replace every occurrence of " $\exists xx \Box_{xx} p$ " in **Standard Contingentism**, then we get:

$$(*) : \exists p \Diamond (\Box p \wedge \Diamond \neg \Box p)$$

(*) is logically equivalent to $\neg \forall p \Box (\Box p \rightarrow \Box \Box p)$, which contradicts the quantified version of the modal axiom 4.

3. Werner's proposal

The above reconstruction of Teitel's puzzle presupposes that $\Box p$ and $\exists xx \Box_{xx} p$ are everywhere intersubstitutable. On such a strong reading of the reduction, **Robustness** (the idea that the reduction of metaphysical necessity to essence necessarily holds, necessarily necessarily holds, and so on) can be understood in a natural way: to say that this reduction is robust just is to say that $\Box p$ and $\exists xx \Box_{xx} p$ are necessarily (necessarily...) identical. This condition is automatically valid in S4, given the necessity of identity.

However, one may naturally wonder what would happen if $\Box p$ and $\exists xx \Box_{xx} p$ are not identified with each other, and---most importantly---how we should understand **Robustness** when we embrace a *weaker* reduction of metaphysical necessity to essence. These are the questions that Werner (2020) wants to answer in his paper.

According to Werner, if we do not identify $\Box p$ with $\exists xx \Box_{xx} p$, then there are two readings of **Robustness**---a *de re* reading, and a *de dicto* reading:

$$\text{De Re Robustness: } \forall p \Box (\Box p \leftrightarrow \exists xx \Box_{xx} p), \forall p \Box \Box (\Box p \leftrightarrow \exists xx \Box_{xx} p), \dots$$

De Dicto Robustness: $\Box \forall p(\Box p \leftrightarrow \exists xx \Box_{xx} p)$, $\Box \Box \forall p(\Box p \leftrightarrow \exists xx \Box_{xx} p)$, ...

It is easy to check that if $\Box p$ and $\exists xx \Box_{xx} p$ are identical, then both **De Re Robustness** and **De Dicto Robustness** are trivially true. However, if $\Box p$ is not identified with $\exists xx \Box_{xx} p$, then there will be a *real* difference between **De Re Robustness** and **De Dicto Robustness**: the former leads to a Teitel-style puzzle, while the latter does not.

To see this, notice that according to **De Re Robustness**, any actually existing proposition p is such that necessarily, $\Box p$ iff $\exists xx \Box_{xx} p$. Then, using "if Socrates exists then he is a human" to instantiate " p ", we get the following statement: $\Box(\Box \text{if Socrates exists then he is a human} \leftrightarrow \exists xx \Box_{xx} \text{if Socrates exists then he is a human})$

Given the distribution axiom K, the axiom 4 and the fact that "if Socrates exists then he is a human" is a necessary truth, we get the following unacceptable consequence: $\Box \exists xx \Box_{xx} \text{if Socrates exists then he is a human}$.

Unlike **De Re Robustness**, **De Dicto Robustness** alone does not generate any Teitel-style puzzle. This is because in **De Dicto Robustness**, the universal quantifier occurs within the scope of the outermost necessity operator. Then, for any actually existing proposition p which is necessarily true but does not necessarily exist, **De Dicto Robustness** wouldn't force us to accept the necessary equivalence between $\Box p$ and $\exists xx \Box_{xx} p$. So we also don't need to accept $\Box \exists xx \Box_{xx} p$. In this sense, if **De Dicto Robustness** is all that we need for a reduction of metaphysical necessity, then Teitel's puzzle is already resolved.

Does **De Dicto Robustness** really suffice to give us a satisfactory reductive account of metaphysical necessity? I do not think so, for mainly two reasons:

1) Firstly, even if **De Dicto Robustness** is true, it is too coarse-grained and can hardly deliver a *sufficient* reductive account of metaphysical necessity to essence. An important point here is that proponents of hyperintensional notions (e.g. grounding, essence, etc.) are usually reluctant to take necessary equivalence to be a sufficient condition for a reduction. If **De Dicto Robustness** is all that we need for a reduction of metaphysical necessity, then there will

also be a reduction of $\Box p$ to $\exists xx \Box_{xx} p \wedge (1+1=2 \vee 1+1 \neq 2)$, which is definitely unacceptable. Viewed from this perspective, Werner's account is at least *incomplete*: even if we accept **De Dicto Robustness**, we still need a reductive relation stronger than *de dicto* modal equivalence.

2) Secondly, Werner's proposal is intelligible only if we are working in a language which contains propositional quantifiers. However, if our language prohibits propositional quantification, then there is no way to make sense of the distinction between **De Re Robustness** and **De Dicto Robustness**. In this situation, the only legitimate reading of **Robustness** which captures the spirit of **De Dicto Robustness** is arguably given by the following schema:

Schematic Robustness: $\Box(\Box\varphi \leftrightarrow \exists xx \Box_{xx} \varphi)$

Unfortunately, **Schematic Robustness** cannot avoid leading to Teitel's puzzle. This is because once we use the sentence "if Socrates exists then he is a human" to instantiate " φ ", we get " $\Box \exists xx \Box_{xx}$ if Socrates exists then he is a human" again.

4. Grounding and contingentism

The fact that *modal equivalence* is too weak for a reduction of metaphysical necessity and the fact that *identity* is too strong jointly suggest that we need a reductive relation which lies in between these two relations. *Grounding* is a perfect candidate.

Following Rosen (2010), I take (partial) ground to be a *relation* that holds between two *facts*. In comparison with the *sentential operator* approach to grounding—which takes grounding to be a sentential operator linking two sentences, the advantage of this *relational* approach is that it allows us to quantify over the *relata* of grounding without introducing any higher-order quantifiers. So, for instance, in our current setting, the idea that grounding is transitive can be regimented through appealing to first-order quantifiers alone:

$\forall x \forall y \forall z ((x \ll y \wedge y \ll z) \rightarrow x \ll z)$ (" \ll " denotes the relation of partial ground)

Since *facts*, serving as the relata of grounding, lie within our domain of quantification, a question naturally arises for contingentists: do facts necessarily exist?

It strikes me that anyone who believes that propositions and properties exist contingently should also believe that facts exist contingently. Moreover, it also strikes me that the existence condition for facts should be structurally similar to the existence condition for properties and propositions. That's to say, we should at least require that if a fact is *about* some individuals, then it won't exist if those individuals don't exist.

Given these *minimal* requirements on the existence of facts, let's consider how we should formulate the reduction of metaphysical necessity to essence in terms of grounding, against the background of contingentism. In the literature, there are already many attempts on this issue that can give us some guidance. Consider the following well-accepted principle, which is first proposed by Rosen (2010):

Naïve Reduction: $\Box\varphi \rightarrow ([\exists xx \Box_{xx} \varphi] \ll [\Box\varphi])$

Here, the result of adding a pair of square brackets "[]" to a formula " φ " (i.e., " $[\varphi]$ ") denotes *the fact that φ* . For the sake of simplification, I will assume that adding square brackets to any closed formula gives rise to a well-formed expression. But this by no measure entails that every closed formula can define a fact that *exists* or a fact that is *true*. So, for instance, given an arbitrarily chosen closed formula φ , we cannot derive either $\Diamond \exists x (x = [\varphi])$ or $\Diamond \mathbf{True}([\varphi])$ (where **True** is a primitive truth predicate, in the sense of Fine (2010)).

Bearing these clarifications in mind, we can now understand the basic idea underlying **Naïve Reduction**. It says that if $\Box\varphi$, then the fact that $\Box\varphi$ is grounded in the fact that it is essential to some objects that φ . Although this principle is widely accepted, it faces two problems if contingentism is true. Firstly, it is directly incompatible with *some* version of contingentism. Consider, for instance, a contingentist who believes that Pegasus does not exist but it possibly exists. Then, this contingentist might also embrace the claim

that necessarily, if Pegasus exists then it is a horse. However, using this necessary truth to instantiate **Naïve Reduction** leads to the following statement:

$$[\exists xx \Box_{xx} \text{if Pegasus exists then it is a horse}] \ll [\Box \text{if Pegasus exists then it is a horse}]$$

Further assuming that grounding is *factive*, we are forced to accept that there are *actually* some entities whose nature is about Pegasus. This yields a puzzle analogous to the one initially proposed by Teitel.

The second problem with **Naïve Reduction** is that it violates **Robustness**—the idea that the reduction of metaphysical necessity to essence should necessarily hold. To see this, let's add a necessity operator to **Naïve Reduction**:

Naïve Robustness: $\Box(\Box\phi \rightarrow ([\exists xx \Box_{xx} \phi] \ll [\Box\phi]))$

Using "if Socrates exists then he is a human" to instantiate " ϕ " in **Naïve Robustness**, we get " $\Box \exists xx \Box_{xx}$ if Socrates exists then he is a human" again.

The above arguments suggest that **Naïve Reduction** should be modified. Here is a plausible suggestion for the modification:

Existential Reduction: $(\Box\phi \wedge \exists x x = [\phi]) \rightarrow ([\exists xx \Box_{xx} \phi] \ll [\Box\phi])$

Existential Reduction says that not all necessary facts are grounded in essentialist facts. Rather, only necessary facts that **exist** can be so grounded. According to this principle, even if the statement that *necessarily, if Pegasus exists then it is a horse* is true, *the fact that necessarily, if Pegasus exists then it is a horse* is not actually grounded in any essentialist fact, simply because this necessary fact about Pegasus does not *exist* at all. Similarly, even if we require that **Existential Reduction** be robust, this will not lead to any Teitel-style puzzle, since *the fact that necessarily, if Socrates exists then he is a human* does not necessarily exist.

Endorsing **Existential Reduction** (and the idea that this reduction is robust) would force us to reject the following comprehension schema:

Modalized Factual Comprehension: $\Box(\phi \rightarrow \exists x x = [\phi])$

This is not a surprising consequence. No faithful contingentist would think that necessarily, any true statement can define a fact that exists, just like no faithful contingentist would think that necessarily, any true statement can define a *proposition* that exists.

However, there is another consequence of **Existential Reduction** that one might find puzzling, namely that if *the fact that necessarily, if Socrates exists then he is a human* is not necessarily grounded in an essentialist fact, then what metaphysical status it should have in a world where it does not exist? Should it be an *ungrounded* fact in that world? If it is ungrounded, then in which sense our reduction of metaphysical necessity to essence can be considered as successful? But if it is not ungrounded, then what fact can be its ground?

It seems that we are faced with a dilemma. On one hand, we do not have a systematic strategy to identify the grounds of this fact in a world where Socrates does not exist. On the other hand, we also don't want to say that this fact could be possibly ungrounded. This will seriously diminish the plausibility of our reductive account. Fortunately, we do have a way of avoiding this dilemma, so long as we endorse the following two principles:

Grounding Entails Existence: $x \ll y \rightarrow (\exists z z=x \wedge \exists z z=y)$

Ungroundedness: $x \text{ is ungrounded} =_{df} (\exists y y=x \wedge \neg \exists y y \ll x)$

Grounding Entails Existence says that if a fact grounds another fact, then both facts exist. **Ungroundedness** says that a fact is ungrounded/fundamental iff it exists and does not have grounds.

Why should we endorse these two principles? First, according to the widely accepted principle of *being constraint*, nothing can stand in any relation (including the grounding relation) if it does not exist. Hence, anyone who endorses *being constraint* should also endorse the above two principles.

Secondly, even if one does not accept *being constraint*, rejecting the above two principles would still lead to absurd consequences. To see this, let's consider the disjunctive fact [Socrates exists or Socrates does not exist], which---according to the standard grounding theory---is grounded in either [Socrates

exists] or [Socrates does not exist]. However, in a possible world where Socrates does not exist, there does not *exist* any fact that can ground this disjunctive fact. Then, people who reject **Ungroundedness** would have to say that in this world, this disjunctive fact is *ungrounded* (since if we do not require that ungrounded facts must exist, then this disjunctive fact would count as ungrounded simply because there is no fact that can ground it). This is absurd. On the other hand, people who reject **Grounding Entails Existence** would have to say that in a world where this disjunctive fact does not exist, it is still grounded in [Socrates does not exist]. Although this claim itself is not completely implausible, it becomes absurd when it is combined with **Ungroundedness**: if being ungrounded entails existence, then being grounded should also entail existence. Then, how can a non-existent disjunctive fact--being neither grounded nor ungrounded--stand in any grounding relation? Since we must pursue a systematic theory of grounding, I take it that the best option for contingentists is to *refuse to endorse any grounding claim which concerns non-existent entities*. This will consequently give us another principled reason to embrace our novel reductive account of metaphysical necessity.

References

- Ditter, A. (2020) "The reduction of necessity to essence", *Mind*, 129(514), 351-380.
- Fine, K. (1994) "Essence and modality: The second philosophical perspectives lecture", *Philosophical perspectives*, 8, 1-16.
- Fine, K. (1995) "The logic of essence", *Journal of Philosophical Logic*, 24(3), 241-273.
- Fine, K. (2010) "Some Puzzles of Ground", *Notre Dame Journal of Formal Logic*, 51 (1):97-118.
- Rosen, G. (2010) "Metaphysical dependence: Grounding and reduction", *Modality: Metaphysics, logic, and epistemology*, 109, 109.
- Teitel, T. (2019) "Contingent existence and the reduction of modality to essence", *Mind*, 128(509), 39-68.
- Sider, T. (2011) *Writing the Book of the World*, OUP Oxford.
- Werner, J. (2021) "Contingent objects, contingent propositions, and essentialism", *Mind*, 130(520), 1283-1294.

The *Tractatus* and Satisfaction

Kevin Yao (Chicago, USA)

Abstract

Wittgenstein's *Tractatus* introduces the notion of a strictly correct method of philosophy that leaves the interlocutor ultimately unsatisfied. Readers of the *Tractatus* such as Cora Diamond, however, have argued that the method of the *Tractatus* itself is to be distinguished from the strictly correct method and further that it is a method which allows for an achievement of a sense of satisfaction of one's philosophical desire. Diamond locates the satisfaction in the process of a 'transformation of desire' wherein the working through to the recognition of a stretch of discourse as nonsense allows the interlocutor to recognize there was nothing at all that could have been what they had wanted to say. I aim to show that while there is room for such an achievement of satisfaction in the methodology of Wittgenstein, one ought not to locate it merely in the recognition of nonsense, which remains ultimately unsatisfying. I introduce language of dissolution and absolution to distinguish between the mere extinguishing of desire and the satisfaction of desire, arguing that the mere dissolution of philosophical desire does not constitute its satisfaction.

The correct method in philosophy would really be to say the following: to say nothing except what can be said, i.e. propositions of natural science—i.e. something that has nothing to do with philosophy—and then, whenever someone else wanted to say something metaphysical, to demonstrate to him that he had failed to give a meaning to certain signs in his propositions. Although it would not be satisfying to the other person—he would not have the feeling that we were teaching him philosophy—*this* method would be the only strictly correct one. (TLP 2009: 6.53)

In this remark of the *Tractatus*, Wittgenstein alludes to a strictly correct method of doing philosophy, one that ultimately leaves its interlocutor unsatisfied. The interlocutor is left only with the recognition that what they thought they had wanted to say was really nonsense—that no meaning had been supplied to the language they thought perfectly meaningful. It may be tempting to take such a description as a sufficient characterization of all that the *Tractatus* itself is doing, or even as all that one could do in philosophy. On such a reading, the injunction to silence in the final line of the book will be taken to reinforce such a notion of the strictly correct method being the complete extent to which one can do philosophy. On this reading, the point of the *Tractatus* then is merely to demonstrate that a collection of apparent thoughts is nothing but nonsense. However, I think another reading of what

Wittgenstein says here can be had, one in which we consider Wittgenstein's turn of phrase in which he speaks of a *strictly* correct method more carefully.

In 6.54, we are told that we are to come to recognize the propositions of the *Tractatus* as nonsensical. However, it is evident to any reader of the *Tractatus* that the book itself does not adhere completely to the strictly correct method: to do so would require the dialogical form of demonstration outlined in 6.53 to somehow be achieved in the monological context of the book. Yet it is important to see here in 6.54 that Wittgenstein's own method in the *Tractatus* need not be exhausted by his playing as one of the two roles allowed in the strictly correct method: either as the interlocutor that unwittingly speaks nonsense or as that of the critic who subsequently demonstrates the nonsensicality of what the interlocutor said. To read the *Tractatus* as simply and only trying to approach the strictly correct method fails to do adequate justice to a further aspect of an activity, one which is performed by the reader of the *Tractatus* in their climbing through and beyond of the propositions of the book. Limiting the reading of the *Tractatus* to just some form of the strictly correct method overlooks a motivation that the author of the *Tractatus* aims to achieve, one which as Wittgenstein writes in the preface, will give "pleasure to a person who read and understood it" (TLP 2009). There is more that the author of the *Tractatus* aims to achieve, and more to the method that he goes about doing so, than would be necessary if *all* that the lines of the *Tractatus* aim to elucidate is their nonsensicality. If the point of the *Tractatus* is merely to provide us with a collection of propositions to be demonstrated to be nonsense and nothing more, then they would elucidate no more than the fact that they are nonsense.

It will be helpful before we get further, however, to first get the strictly correct method, as Wittgenstein thought of it, into clear view. Here, the issue I want to emphasize is the way in which, even under the strictly correct method, the methodology is still one of demonstration. We can get at what is meant here by demonstration if we consider James Conant and Cora Diamond's criticism of a reading of that method offered by Peter Sullivan. Sullivan takes the strictly correct method to be one in which the *Tractatus* provides its reader with all that they need to recognize the deficiencies of sense in stretches of discourse. The philosopher is able to recognize and then point out stretches of nonsense as they are said. The method is one of using the *Tractatus* to pass verdicts on

discourse, from which the interlocutor is able to be told of the nonsensicality of what they say. Conant and Diamond rightfully point out that such a reading of the method is “in any case no *demonstration* that the person has come out with nonsense” (Conant and Diamond 2005: 73, my emphasis). The point being made here is that Sullivan’s account is not even one in which the interlocutor is led to come to see that what they have said is nonsense—they are simply told so. It is a misunderstanding of Wittgenstein’s methodology to take one as capable of coming to see the nonsensicality of what they say prior to an engaged demonstration. The speaker needs to come to see that what they had said had not been supplied with meaning, nor could there have been anything that would have sufficed as what they wanted to say.

I raise this criticism of Sullivan because such an account “would indeed be dissatisfying” (Conant and Diamond 2005: 73), but this is not at all the sort of dissatisfaction we should attribute to the strictly correct method understood properly. The dissatisfaction brought about by Sullivan’s reading is one in which the interlocutor is told that they are speaking nonsense, but have yet to come to recognize through demonstration *that they are in fact speaking nonsense*. The dissatisfaction of the strictly correct method, on the other hand, is felt by someone who has come to recognize they have been speaking nonsense, but has not yet extinguished the philosophical desire to speak what the interlocutor now sees as nothing at all.

Let us now turn back to the distinction between the strictly correct method and the method of the *Tractatus*. It is important to see that the critical difference between what is outlined in 6.53 and what the *Tractatus* takes itself to be doing in 6.54 is not that 6.54 stands as the next best thing after 6.53: that Wittgenstein would have simply followed the strictly correct method had it been possible to do so in a book. Instead, the difference is that of satisfaction: the *Tractatus* is going about in a way beyond the strictly correct method in order to engage the reader in an activity that will satisfy their philosophical desire. If we read the *Tractatus* as making a deliberate deviation from the strictly correct method for just such an ambition, then we ought to read Wittgenstein’s use of the word ‘strictly’ with some irony. Although the method outlined in 6.53 might be the one that is *strictly* correct, it is not, in fact, all that is being employed by the *Tractatus* to achieve satisfaction.

This, I think, leaves room in the *Tractatus* for one to achieve a sense of satisfaction of their philosophical desire unavailable to one who adheres to the strictly correct method. The strictly correct method only gets us to see that there is nothing that is what we want to say, but even as a method of demonstration it does not satisfy the desire to speak in that way. The *Tractatus*, of course, does not aim to satisfy us by giving us some means by which we *can* come to speak nonsense, but instead leads the reader through a transformation of desire, wherein the philosophical desire can in the end nonetheless be said to have been satisfied in a way different to what was initially envisioned.

I want to now turn to Diamond's conception of the transformation of desire, where an initial desire to speak in a certain way is left satisfied in light of the crumbling of that view into nonsense. In particular, I want to examine Diamond's treatment of solipsism, where she takes herself to be following the Tractarian way of going about. In Diamond's account, we follow a crumbling of a view of solipsism, such that the solipsist comes to see that nothing could have been what they had wanted to say with the language of solipsism. Diamond locates the satisfaction in the crumbling of solipsism, but I think we will have to see that if we want to hang on to the notion of achieving satisfaction in the *Tractatus*, it cannot be here as Diamond locates it.

The view of solipsism we are concerned with rejects Russell's two-limit view: "the limits of the objects which I can directly *name*, are narrower than the limits of the world" (Kremer 2017: 65). Russell claims that there are things that exist in the world that I can nevertheless not assign a name to—they exist, so to speak, *outside* of my limit of naming. The solipsist instead posits a single limit, wherein the limit of that which I can name *is* the limit of the world: 'my world' is '*the* world'. But the solipsist is then faced with maintaining the coherence of the one limit view. Diamond argues they will find themselves falling into nonsense:

[The solipsist] draws on ordinary ways of using 'the' in contrast with 'my'/'me'/'mine' in identifying something, but the ordinary ways of making such contrasts are lacking, and no other mode of making them has been

provided. So the final stage of the collapse of the solipsist's attempt to say his solipsism involves a recognition that no saying of anything would be what he wants. (Diamond 2012: 19)

In giving up the two-limit distinction, the solipsist is also giving up on the distinction in language necessary for them to say their rejection. If 'my world is *the* world', then the sentence means nothing, because it no longer identifies two things. The solipsist recognizes that they have fallen into nonsense, but critically, this is not an affirmation of the two-limit distinction. Diamond takes Wittgenstein's reference of the "extent to which solipsism is a truth" (TLP 2009: 5.62) as how solipsism is a turn away from the two-limit view and a rejection of the desire to get beyond the limit of what I can name. But if we speak about 'the inability of getting beyond the limits of our language' then it seems like we are positing a limit, with things on the other side of that limit. The result of falling into nonsense is rather the giving up of the idea of a limit itself; giving up that there is something that we cannot name. There is thus no return to a two-limit view. The idea of a limit is abandoned as one sees that there is nothing beyond what one can say; that there is no limit dividing the sayable and unsayable because there is nothing unsayable.

Diamond is keen to dispel the notion that recognition of the philosophical limits of what we can say ought to lead to a sense of resignation, held in contrast to satisfaction. An important notion underlying this account is that of a non-contrastive limit, wherein the limit is not drawn between sayable and unsayable things, but between what is sayable and nothing at all. Diamond makes the case that person who recognizes the crumbling of their thoughts need not be "resigned to the impossibility of saying something that he wants to say" (Diamond 2012: 20). It is not the case that they come to recognize what they had wanted to say was unsayable; rather the clarity that the interlocutor reaches is simply that *nothing* could have been what they had wanted to say.

By understanding Wittgenstein like this, we can begin to get a better sense of the injunction to "throw away the ladder" in 6.54. Wittgenstein speaks of the "extent to which solipsism is a truth" which, at first, leads us look for some fact in the world that might serve as *that* truth. Yet in thoroughly examining solipsism itself, we see that it crumbles as a distinct view. But in working *through* how it crumbles, the misunderstanding that the two-limit distinction

offers is cleared away. Solipsism is a ladder we must climb, but once we have climbed it, we realize we must throw *it* too away.

I have only given a very rough account of the crumbling of solipsism, but the methodological point I want to emphasize is how Diamond takes the collapse as a process in which we are ultimately left with, not something unsayable, but a recognition that there was nothing at all we wanted to say. This, I think, is right, although I will not argue it here. What I want to instead examine is the conclusion Diamond gives to the discussion on solipsism:

That one moves through and beyond saying such things as ‘The world is my world’ or ‘The subject is a limit of the world’ is what enables the collapse of solipsism as a distinct view to be a satisfaction of the desire to speak solipsism. (...) What I am here taking for granted is that a theme of all of Wittgenstein’s philosophy is the transformation of philosophical desire, i.e. the idea that such desire can be understood to be satisfied, at the end of the activity of philosophy, by something different from what one had originally taken oneself to want. (Diamond 2012: 20)

Diamond identifies a moment of satisfaction that follows from the collapse of the nonsensical view, in contrast to the resignation that would follow if recognizing that one has fallen into nonsense means that there is something they are unable to say. But I think to identify satisfaction here is to place it too early. It is not at all clear, once we have witnessed the collapse of a view, what the philosophical desire is anymore. Diamond wants to say that there is a transformation of desire, where the philosophical desire to speak something is satisfied by seeing how there was nothing at all we wanted to say. But the collapse of solipsism as a distinct view is also the collapse of it as something one might be so desired to speak of. A notion I am here holding is a strong link between the desire of speak something and the robustness of what one wants to speak. A view only stands as attractive to you if it can stand as something meaningful. The desire to speak solipsism, for example, only holds so long as solipsism seems to you as a distinct and robust view—that is, as a meaningful stretch of language. Once that robustness has collapsed, as it does under Diamond’s account, whatever was the thing the interlocutor desired is no

longer in view. This is not to say one could not find themselves once again desiring to speak solipsism, but that such a desire only returns just as one once again comes to see solipsism as a meaningful view of which one can speak of.

The account of satisfaction that Diamond locates here ought to echo that of the strictly correct method. It is precisely the method that is unsatisfying because it only offers the person a clear view that what they were trying to say is nonsense. Diamond is right to say that this does not leave us resigned within the limits of language (taken contrastively), but we should not take this lack of resignation to be a sign of satisfaction. We should distinguish the mere extinguishing of desire from its satisfaction. That one feels no desire to speak in a certain way does not mean that that desire has in turn been satisfied, because *they may now not feel the force of that desire at all*. What I think is leading Diamond's conception of the transformation of desire to locate the satisfaction in the crumbling of of a view is a muddling up of the purely negative notion of the extinguishing of a philosophical desire with its being satisfied. A lack of resignation is the commonality between the two, but one cannot locate the positive sense of satisfaction in just the purely negative absence of the desire.

Consider the contrast between the notions of what I will call absolution and dissolution. The term 'absolution' is concerned with the freeing of one from obligation, but it does so by way of a very interest move, distinct from terms like 'dissolution' or 'resolution', even if the end results pertaining to the obligation are all the same. To absolve one of their obligations is neither to deny the obligation nor that one was under such obligation: to absolve someone of being guilty of theft is not to deny that theft is a crime. It is recognition that such obligation no longer applies. To resolve, on the other hand, is to accept the obligation, for which we then supply an answer or response to satisfy its demands. To dissolve, finally, gets out from being under the obligation by doing away with the obligation entirely. The way I propose to use these terms pertains to desire and its satisfaction. To resolve is in some way to satisfy the philosophical desire, not through a transformation of desire, but in how the interlocutor initially intended to go about it. This is far from the sort of satisfaction that Wittgenstein is concerned with. Diamond, on the other hand, is presenting a case of dissolution—one in which the collapse of solipsism leads to the collapse of the desire to speak it—but passes it off as that of

absolution: genuine satisfaction of that desire. With these two moves distinguished, it becomes clearer that Diamond's account of dissolution cannot be one of satisfaction, precisely because it leads to the dissolution of the desire that needs to be satisfied. The collapse of solipsism cannot be the satisfaction of the desire to speak solipsism, because the standing of solipsism as a distinct view is what makes possible the desire in the first place.

References

Conant, James; Diamond, Cora (2005) "On Reading The *Tractatus* Resolutely: Reply to Meredith Williams and Peter Sullivan", in: Max Köbel and Bernhard Weiss (eds.) *Wittgenstein's Lasting Significance*, London: Routledge, 42–97.

Diamond, Cora (2012) "The *Tractatus* and the Limits of Sense", *The Oxford Handbook of Wittgenstein*, 240–75.

Kremer, Michael (2017) "To What Extent Is Solipsism a Truth?", *Post-Analytic Tractatus*, 59–84.

Wittgenstein, Ludwig (2009) *Tractatus logico-philosophicus*, revised edn., London: Routledge & Kegan Paul [TLP].

Coinciding Groups: A Higher-Level Pluralist Approach

Yiwen Zhan (Beijing, China)

Abstract

I defend a ‘pluralist’ account of groups, according to which a group is neither a (singular) mereological sum nor a set, but a plurality of its members. In particular, I will offer a novel semantic analysis that treats groups as higher-level pluralities (i.e. pluralities of pluralities, etc.). This analysis, while being able to neatly accommodate the (metaphysician’s) requirement that group-formation should be ontologically innocent, also meets the desiderata for capturing the critical features of groups. Moreover, we will even be able to deal with sensitivities not only to ways of compositions but also to ways of constitutions, and in particular, to offer a way to differentiate between coinciding groups.

1. Features of Groups

Groups are ubiquitous. In Landman (1989), Fred Landman has offered an elegant and vivid depiction:

We find ourselves somewhere in ‘Das Kapital’. *The Farmers* and *The City-proletarians* have been introduced, and they clearly are type-1 groups ... Now *The Working Classes* are introduced: they consist of the farmers and the city-proletarians, but again *The Working Classes* is treated as a group that does not inherit the properties that both the farmers and the city-proletarians as groups have, so *The Working Classes* is a type-2 group here. But, of course, one page later *The State* is introduced as consisting of *The Working Classes* and *The Exploiting Classes*, and again it does not inherit their properties, and hence is a type 3-group. But there are states, and a couple of pages later we meet *The States of Central Europe* and *the States of Western Europe* who distrust each other, which is the sum of two type 4-groups, etc.

Yet these intuitions give rise to many metaphysical questions: How can we identify a group? Under what conditions can a group be said to be the same with another? Groups are generally deemed as having members. Yet a group may stay the same while having different members at different circumstances. What’s more: two different groups may actually have exactly the same members but stay different—the so-called problem of “coinciding groups”.

In general, we may list four main desiderata for a theory of groups:

- (i) *The membership-compositional structure* of groups (e.g. subgroups, iterability, non-transitivity, etc.);
- (ii) *Contingent (non-rigid) membership*: e.g. ‘the US Supreme Court’ (cf. Uzquiano (2004));
- (iii) The existence of *coinciding groups*: differentiating ‘*a qua F*’ from ‘*a qua G*’; and
- (iv) *The ‘structural’ differences* between groups like queues and groups like com- mittees.

These features indicate that the metaphysical structure of groups is rather fine-grained. For one thing, feature (i) indicates that the compositional structure of groups is strictly more fine-grained than standard extensional mereology. Therefore, we can regard the metaphysics of groups as a generalization of standard mereology. In other words, any composite objects can be seen as groups. But for ‘full-fledged’ groups (like social groups), they are not only sensitive to the structure of mereology and composition, but also sensitive to their ways of ‘constitution’.

In this light, I will assume that both (iii) and (iv) concern the same problem of group *constitution*. Namely, different coinciding groups are different just in case they have different internal structures among their members. Of course, this will not solve all the problems involving coinciding groups. For instance, if we adopt the view that any social establishment is a group, then a group might have zero members. But two coinciding empty groups arguably have no internal structural difference. In this paper, however, we will assume that all groups are non-empty. Thus, according to our approach, we do not require a one-to-one correspondence between groups and social establishments. A same group, for instance, may correspond to two different social establishments. Social establishments in this sense are even more fine-grained than groups. But will restrict our attentions to groups understood as non-empty entities whose sameness condition is fine-grained to the level, and only to the level, of sensitivity to internal constitutions. It might seem natural to view conditions (ii), (iii), and (iv) as all involving modality.

However, modality cannot do all the tricks. Indeed, we can ‘simplify’ the modal features of (ii) by equating groups with world-time-slices or ‘stages’ (Wilhelm, 2022). But then, the same modal trick won’t work for solving (iii) and (iv): True, we can say that ‘*a qua F*’ and ‘*a qua G*’ have different modal and temporal parts. But if we want to avoid circularity with respect to explanations of contingent membership, we are no longer be able to disambiguate coinciding groups, unless with certain extra ontological commitment to things like ‘modal parts’ (Wallace, 2019) or hylomorphic ‘formal parts’ (Fine, 1999; Uzquiano, 2018). However, we will later show, we can meet the four desiderata without paying any of such extra ontological costs.

2. Singularist Accounts and Their Limits

In formal semantics, groups are typically treated as singular entities. There are both technical conveniences and linguistic intuitions in favor of singularism. It thus seems natural to take standard group nouns (such as ‘pile’, ‘deck’, ‘team’, ‘committee’, and indeed ‘group’) as denoting groups qua singular entities. These group nouns are indeed morphologically singular, and they help to form countable classifier phrases.

But these linguistic data are not yet decisive. (See Oliver and Smiley (2016, §15.1) for more discussions.) Moreover, the singularist accounts are not even technically satisfying. Consider the most straightforward principle of group-formation.

(Formation) $\forall xx \exists y. \text{group}(y) \ \& \ y = g(xx)$

It says nothing about the structure of a group, neither the criterion for identifying two groups. Of course, we can add a principle that says that two pluralities always form the same group if they have the same members:

(Identity) $g(xx) = g(yy) \leftrightarrow xx \approx yy$

But this already defies *Contingent Membership*. And more seriously, it defies *Plural Cantor*, i.e. that there are more pluralities than singular objects.

A more principled and more structural account might treat groups as mereological sums. Based on Classical Extensional Mereology, for any plurality

of entities such that something is one of them if and only if it is identical to either a_1, a_2, \dots or a_n , we can write their (unique) sum as $a_1 \sqcup a_2 \sqcup \dots \sqcup a_n$, or equally, as $\sqcup aa$. Identifying groups with sums would then yield $g(xx) = \sqcup xx$. But as mentioned earlier, mereological sums are still too coarse-grained for groups. For one thing, it still defies *Contingent Membership*. Also, more generally, the compositional structure of groups (subgroups, iterability, non-transitivity) is also more fine-grained than classical mereological structures. Consider the following Wall–Brick–Atom scenario where a_1 and a_2 compose b_1 , a_1 and a_3 compose b_2 , a_2 and a_3 compose b_3 , and b_1, b_2, b_3 together compose c . Since the parthood relation results in a complete atomic join semilattice, we have $\sqcup aa = \sqcup bb$. But in general, we have $xx \approx yy \rightarrow \sqcup xx = \sqcup yy$ but not vice versa. In fact, if we regard plural identity ‘ \approx ’ as regular identity, it can be shown that the axioms of Classical Extensional Mereology jointly entail mereological nihilism, i.e. that there are no composite objects like b_1, b_2 , or c (Sider, 2013).

The lesson is that any singularist account of groups must be sufficiently fine-grained so that it can smoothly distinguish between groups that occupy the same (mereological) portion of reality. To this end, Snyder and Shapiro (2022) proposed a different account of groups, according to which groups can be treated semantically as (roughly speaking) sets of sums. Use the Wall–Brick–Atom case again: while the sum of b_1 and b_2 is equivalent to the sum of $a_1 \sqcup a_2 \sqcup a_3$, to retrieve group-iteration, Snyder and Shapiro defined a group formation operation ‘ g ’, such that we can form two groups $g(a_1 \sqcup a_2)$ and $g(a_1 \sqcup a_3)$, which, in turn, can form a new, higher-level group $g(g(a_1 \sqcup a_2) \sqcup g(a_1 \sqcup a_3))$, which is different from $g(a_1 \sqcup a_2 \sqcup a_3)$. Such account enables the iterative intuition of group-membership and hence is sufficiently fine-grained. It enables analyses of collective properties at different levels, e.g. “This committee and that committee control each other”; as well as count predicates e.g. “Those are two committees.”

However, there are still limitations of Snyder and Shapiro’s approach: First, they assume that for every sum, there is a unique group, and vice versa, but then it fails to account for coinciding groups. Second, while their approach enables group-iteration, at each level of group iteration, it still lacks

explanation of how the members of a newly formed group do not get to ‘forget’ their own members at lower levels. The only available explanation seems to treat groups simply as sets. In fact, structurally their analysis of groups is akin to the analysis of sets. But seeing groups as sets requires commitments to abstract entities and hence are ontologically expensive. Finally, the set analogy reveals that their account even has the group-version of Russell’s Paradox: Assuming unrestricted group formation and arbitrary sums:

(SG-Formation) $\forall x \exists y. \text{group}(y) \ \& \ y = g(x)$

(S-Comp-A) $\exists x. \varphi_x \ \& \ \text{atom}(x) \rightarrow \exists y \forall x (\text{atom}(x) \rightarrow (xPy \leftrightarrow \varphi_x))$

To avoid the paradox, they claim that group formation is only potential:

(SG-Formation-P) $\Box \forall x \exists y. \text{group}(y) \ \& \ y = g(x)$

3. The Pluralist Interpretation of Groups

In light of these, I submit that we have better reasons to adopt a pluralist interpretation. Following Oliver and Smiley (2016), we should regard groups nouns as ‘pseudo- singular terms’: they are morphologically singular, but semantically plural. (Also cf. Horden and Lo’pez de Sa (2020)). Consider e.g. the simplest kind of pluralist approach, according to which a group is just a plurality of entities. Such an approach immediately has the benefit of ontological innocence: i.e., since it commits to nothing over and above the entities that the plural quantifier ranges over, plurality-formation is an ontological free lunch, just like what the (problem-plagued) principle of ‘Composition As Identity’ in mereology wants to achieve (Loss, forthcoming). Moreover, the pluralist account of groups also requires no extra existence conditions for some entities to form a group. Finally, the pluralist account does not have to worry about Russell’s Paradox.

But such a simple pluralist interpretation also has its drawbacks: Above all, under such interpretation, groups are not really iterative (hence no genuine presentation of ‘groups of groups’ available). One way to overcome this is to mimic Snyder and Shapiro (2022) and allow the plural quantifier to range over mereological sums. With the help of extensional mereology, the simple pluralist account can thereby derive correct truth conditions for collective

predicates at different ‘levels’. For example, consider two groups: the (group of) bricks and the (group of) atoms, we are now able to differentiate them as two different groups—If we use $[x_1, \dots, x_n]$ as short for the plural definite description of the plurality consisting precisely of entities x_1, \dots, x_n , then we have $[b_1, b_2, b_3] [a_1, a_2, a_3]$. And this generates narrower contexts where $\text{brick}([b_1, b_2, b_3])$ is true but $\text{brick}([a_1, a_2, a_3])$ is false.

Even so, the simple pluralist interpretation is still not fine-grained enough: it falls short of delivering the iterative structures of groups, and the distributive behaviors of their properties thereof. In fact, a closer inspection into the distributivity of a group’s properties is needed not only for its own sake, but is also crucial for tackling the problem of coinciding groups, e.g.:

- (1) The cyclist club and the chess club coincide.
- (2) The Energy Committee and the Budget Committee coincide.

It is common to distinguish between *distributive* and *collective* properties. The former include predicates like *smile*, *sing*, *wear a dress*, as well as predicative expressions governed by overt distributive quantifiers like *each*. The latter typically include predicates like *gather*, *meet*, *scatter*, *surround the fort*, etc.

Distributive predicates cannot help us differentiate between coinciding groups. Suppose both the cyclist club and the chess club consist of just three persons: Amy, Ben, and Charlie. And suppose the cyclist club has a distributive property—e.g. that they sing. This then entails that Amy sings. But since Amy cannot both be singing and not singing, it can’t be that the chess club does not have the distributive property of singing. However, arguably, collective predicates like *meet* cannot differentiate between coinciding groups either: in a less narrow context, whenever the cyclist club is said to be meeting, there is good sense to take the statement that the chess club is meeting as equally true. (At least it does not force one to judge the latter as false, just like you cannot force one to accept linguistic opacity.)

Nevertheless, I argue that there are *some* collective properties such that, when predicated of two different (yet coinciding) groups, it may force one to judge one as true but another as false. These properties essentially invoke the internal structures of different groups. Consider:

(3) The cyclist club has strong female players.

(4) The chess club doesn't have strong female players.

Intuitively, (3) and (4) can be both true. In this context, arguably, *having strong female players* is a genuinely discerning collective property had by the cyclist club but not by the chess club. And this is precisely because the property is able to discern the asymmetrical internal structure of each of the coinciding groups.

To this end, we need to distinguish between properties that are *homogeneous* and those that are *non-homogeneous*. While all distributive properties are automatically homogeneous, collective properties can be either homogeneous or non-homogeneous. Predicates like *gather* and *scatter* are homogeneous, while *having strong female players*, just like *form a pyramid*, *constitute a majority*, *be denser in the middle*, are non-homogeneous.

A thorough inspection of homogeneity is beyond the scope of this paper. But it suffices for current purpose to adopt the following analysis: if a collective property is said to be had by a group homogeneously, it means there the group can be divided into subgroups down to a *certain* granularity level α , such that at that level, each subgroup has that property, too (cf. Champollion (2017)). Viewed in this way, the difference between distributive and homogeneous collective properties are simply relative to how deep down the group admits of being divided into subgroups.

What about non-homogenous properties? They cannot be said to be distributive at any lower level α of division of the original group. Suppose there are three atoms *a*, *b*, and *c*, we will cash out different levels of granularity in terms of higher-level pluralities. Thus, [*a*, *b*, *c*] is the level-1 plurality that includes all the atoms, while [*a*, *b*, *c*, [*a*, *b*], [*b*, *c*], [*a*, *c*], [*a*, *b*, *c*]] is

the level-2 plurality that includes everything (cumulatively) at granularity level-1. Suppose b is Ben and c is Charlie, then $[b, c]$ doesn't share the property of *having strong female players*, thus the property is non-homogeneous.

However, we can utilize conceptual covers to restrict the membership of a higher-level plurality. For instance, in the example above, $m = [a, b, c, [a, b], [b, c], [a, c], [a, b, c]]$ is the fully cumulatively inclusive plurality of level 2. But if we apply a certain cover C on it, we might exclude certain lower level pluralities from m , thus obtaining $m_C = [[a, b], [a, c], [a, b, c]]$. If so, the property *having strong female players*, while originally non-homogeneous and hence non-distributive for m , now is distributive for m_C . In general, the conceptual cover C can be identified by finding out the *symmetries* in m with respect to the property *having strong female players*, as if m_C is a 'quotient' of that property, such that elements like $[b, c]$ is filtered out. With this technique, we are then able to turn a non-homogeneous property into a homogeneous one under a conceptual cover.

Putting all these together, we can now reduce all kinds of properties—collective or non-collective, homogeneous or non-homogeneous—to the behaviors of distributivity of the said group. For this, we need to not only (i) find a contextually salient level α of granularity to divide the group into (possibly overlapping) smaller parts, but also (ii) apply a contextually salient cover C_α to make sure that each part under that division shares that property.

4. Groups as Higher-Level Pluralities

By treating groups as higher-level pluralities, we will be able to regard two coinciding groups (which have different constitutions) as having their members playing different (*internal*) roles.

The idea of internal roles shares much similarity with the so-called 'positionalism' in the literature on neutral relations (see e.g. Fine (2000); Dorr (2004); Dixon (2018)). While each relatum in a positionalist account occupies a particular 'position' or 'slot', this essentially involves treating the relata as forming a *list*. It is well noted in the studies of plural logic that lists can be sensitive to *orders* and *repetitions*, e.g.:

(5) Amy and Ben completed the test, *in that order*.

With the structure of higher-level pluralities at hand, it is easy to reduce lists to higher-level pluralities: for an indexed list $\langle a_1, a_2, a_3, \dots, a_n \rangle$ of n (possibly repetitive) items, we can accordingly generate a plurality $n - 1$ level higher: $[a_1, [a_1, a_2], [[a_1, a_2], a_3], \dots, [[a_1, a_2], a_3], \dots, a_n]$ This lends support of treating groups as higher-level pluralities.

To identify distinct but coinciding groups, we need to look into non-homogeneous properties that are able to tell the coinciding groups apart. For each coinciding group, e.g. the cyclist club, it has a set of *canonical differentiating properties*. The conjunction of these properties then give rise to a unique conceptual cover C_α that, when applied to the level α plurality under C , everything that can be truthfully said of the cyclist club is distributive at level α . In sum, we can then treat coinciding groups as each corresponding to a different *higher-level plural definite description*, while their differences are set by the different conceptual covers, which, in turn, is determined by each of the coinciding groups' canonical differentiating properties.

In conclusion, as a pluralist approach to groups, our account enjoys all the main virtues of pluralism: the metaphysics of groups is rigid and enables unrestricted composition. Moreover, viewing groups as higher-level pluralities (or more precisely, as higher-level plural definite descriptions) can solve the problem of coinciding groups. It thus nicely captures different groups' internal structures and their ways of constitutions. While we can regard the higher-level constructions as inherently a conceptual construct (compare to cover semantics or cluster semantics, etc.), groups in our account thus remain ontologically innocent. The membership-compositional structure of groups (i.e. subgroups, non-transitivity, etc.) is also straightforwardly retained.

References

- Champollion, Lucas (2017). *Parts of a Whole: Distributivity as a Bridge Between Aspect and Measurement*. New York, NY: Oxford University Press.
- Dixon, T. Scott (2018) "Plural Slot Theory", *Oxford Studies in Metaphysics* 11, 193–223.
- Dorr, Cian (2004) "Non-symmetric relations", *Oxford Studies in Metaphysics* 1, 155–92.
- Fine, Kit (1999) "Things and Their Parts", *Midwest Studies in Philosophy* 23(1), 61–74.

Fine, Kit (2000) "Neutral Relations", *The Philosophical Review* 109, 1–33.

Horden, John and Dan López de Sa (2020) "Groups as pluralities", *Synthese* 198(11), 10237–10271.

Landman, F. (1989 "Groups, I", *Linguistics and Philosophy*, 12(5), 559–605.

Loss, Roberto (forthcoming) "Composition as Identity and the Innocence of Mereology", *Philosophy and Phenomenological Research*.

Oliver, Alex and Timothy Smiley (2016) *Plural Logic*. Second Edition. Oxford University Press.

Sider, Theodore (2013) "Against Parthood", *Oxford Studies in Metaphysics* 8, 237–293.

Snyder, Eric and Stewart Shapiro (2022) "Groups, sets, and paradox", *Linguistics and Philosophy* 45(6), 1277–1313.

Uzquiano, Gabriel (2004) "The supreme court and the supreme court justices: A meta-physical puzzle", *Noûs* 38(1), 135–153.

Uzquiano, Gabriel (2018) "Groups: Toward a Theory of Plural Embodiment", *Journal of Philosophy* 115 (8), 423–452.

Wallace, M. (2019) "The Lump Sum: A Theory of Modal Parts", *Philosophical Papers*, 48(3), 403–435.

Wilhelm, Isaac (2022) "Pluralities, counterparts, and groups", *Philosophical Studies* 179(7), 2133–2153.

Acknowledgement

Work on this chapter received support from the Major Project of National Social Science Fund of China (21&ZD049) and the Major Project of the Key Research Base of Humanities and Social Sciences of the Ministry of Education of China (22JJD720007).

A Wittgensteinian Survey of Ethical Boundaries

Muhammad Zulnoorain (Vienna, Austria)

Abstract

This paper discusses the application of Wittgenstein's notion of "family resemblance" to the problem of defining ethical concepts. It offers a demonstration of a heuristic suggestion made by Wittgenstein in §77 of the *Philosophical Investigations* that would allow for the achievement of an "aspect shift"; one from finding ethical concepts to be arbitrary and vacuous when failing to sharply define them, to seeing them as having *some* definition or boundaries, albeit, variable and complex ones. The demonstration takes the form of a Wittgensteinian "survey" of ethical examples that reveals three kinds of defining relations operative in the use of ethical concepts: 1) direct similarities between ethical examples, 2) inter-relationships between ethical language-games, and 3) shared ethical customs in a "form of life". It concludes that contrary to a popular objection, family resemblance concepts, including ethical ones, are resistant to the charge of vacuity and arbitrariness directed against them, just as Wittgenstein thought.

1. Introduction

This paper discusses the application of Wittgenstein's notion of "family resemblance" to the problem of defining ethical concepts; one that differs from existing accounts on this topic (Klagge 2018) (De Mesel 2019) (Kuusela 2020). It involves a demonstration of a heuristic suggestion made by Wittgenstein in *Philosophical Investigations*:

But if the colours in the original shade into one another without a hint of any boundary, won't it become a hopeless task to draw a sharp picture corresponding to the blurred one? Won't you then have to say: "Here I might just as well draw a circle as a rectangle or a heart, for all the colours merge. "Anything and nothing is right." — And this is the position in which, for example, someone finds himself in ethics or aesthetics when he looks for definitions that correspond to our concepts. In this sort of predicament, always ask yourself: How did we *learn* the meaning of this word ("good", for instance)? From what sort of examples? In what language-games? Then it will be easier for you to see that the word must have a family of meanings. (PI 2009: §77)

Wittgenstein is saying that if you look for a "sharp" (essential) definition of certain concepts, e.g. ethical concepts such as "good", you will conclude that "anything - and - nothing is right" i.e., the concepts will appear *arbitrary* and

vacuous. As a solution, he suggests that if we ask ourselves about the various examples and language-games in which such concepts are learned, if we “don’t think, but look!” (PI 2009: §66), it will become “easier to see” that they form a “family of meanings” i.e., belong to the same general concept. We achieve an “aspect shift” from *thinking* that ethical concepts have no definition or boundaries to *seeing* that they have some definition or boundaries, even if these can’t be sharply drawn.

2. Examples in a “Similar” Language-Game

The first kind of relations we might see between ethical examples are direct similarities between the actions being “modified” by the use of an ethical term. By virtue of these similarities, our use of ethical (or aesthetic) terms for certain actions constitutes a *similar* language-game:

We do not as children discover the quality of beauty or ugliness in a *face* and find that these are qualities a *tree* has in common with it. The words ‘beautiful’ and ‘ugly’ are bound up with the words they modify, and when applied to a face are not the same as when applied to flowers and trees. We have in the latter a similar ‘game’...and similarly in ethics: the meaning of the word ‘good’ is bound up with the act it modifies. (AWL 1979: 35)

Wittgenstein is saying that our use of the words “beautiful” or “ugly” to describe *flowers/trees*, as opposed to describing *faces*, constitutes altogether *different* language-games because the meaning of the modifying words (“beautiful” or “ugly”) varies (“are not the same”) because of the difference in characteristics between the objects being modified by them (“faces” versus “flowers” or “trees”). Describing a face as beautiful involves referring to characteristics such as the eyes, nose, smile etc., and the experience of a particular aspect of beauty (e.g. physical attraction). By contrast, describing flowers or trees as beautiful involves referring to characteristics such as the trunk, bark, leaves, petals, color-patterns, fragrance etc. and the experience of their beauty might be something like the serene effect experienced in being outdoors. We recognize this difference when we read the statements:

My friend thinks Mark has a beautiful smile.

As he strolled in the park, Mark appreciated the beautiful fall foliage.

Hence, we are in a *similar* language-game when there are direct similarities between the descriptions of actions/objects being modified by ethical/aesthetic terms, by virtue of which we notice a *similar aspect*, of goodness or beauty, in them. E.g. actions such as “murder”, “fighting”, “assault” etc. are similar in that they are acts of “violence” and “harm”. By virtue of this similarity between them, we are taught to use words like “bad” or “wrong” for these actions in what Wittgenstein calls “an ethical sense” (PG 1974: 77), as opposed to the sense in which we use “bad” for “guitar-playing” or “wrong” for “calculating”. We may view such similarities as what Wittgenstein called similarities “in the small” (PI 2009: §66).

3. Inter-related Language-Games in a “Family” of Games

The second kind of relation between ethical examples lies in the “combination of a very large number of inter-related games” which make “a single concept” (PG 1974: 77). This makes ethical concepts *polythetic*, not *polysemic*, i.e., they *do not* have “four or five different meanings” (AWL 1979: 33). What makes them single concepts are inter-relationships described by Wittgenstein as:

- “continuous transitions” between members of the concept. (AWL 1979: 33)
- “overlapping” and “sharing of features” between members of the concept. (PG 1974: 75)
- “connections”, “relationships” between “facet[s] of the use” of the concept. (PG 1974: 77)
- “a complicated network of similarities, overlapping and criss-crossing” (PI 2009: §66)

We may understand such “similarities in the large” as those between the concepts *surrounding* the descriptions of various actions in various ethical language-games. They can be seen, as Baker and Hacker put it, by looking “*around* the phenomena” as opposed to looking “*at*” the phenomena (Baker and

Hacker 2004: 221). They involve the sort of relations that require “seeing connections” operative in the use of our words by “finding and inventing *intermediate links*” (PI 2009: §122). Wittgenstein writes, “It [the concept] may connect the objects like the links of a chain, so that one is linked to another by *intermediary links*”. (PG 1974: 75)

Seen this way, we will notice two kinds of inter-relationships between ethical language-games. One inter-relationship is that involving *the use of the same ethical term for a very large number of different actions*. For example, when we teach a child that “lying is wrong” and “stealing is wrong”, we want the child to see that both actions as ethically similar, as “bad” or “wrong”, *despite* differences between their descriptions *qua* mere actions. In terms of their descriptions as mere actions, lying (“to utter a falsehood”) and stealing (“to take something that belongs to someone else”) are different. However, we can find or invent intermediary conceptual links between their descriptions whereby both lying and stealing are similar in that they involve “depriving” another person of something; of the truth (in case of lying), or of a possession (in case of stealing). Using this method, we can even see ethical relations between more “distant” examples that “belong to the same family without any longer having anything in common” (PG 1974: 75) e.g. between “abortion” and “tax evasion”, or between “murder” and “environmental pollution”.

The second kind of inter-relationship between ethical language-games is one involving *the use of a different ethical term for the same action*. The difference in use is due to variations to a “minimal” description of the action, and by extension, to the concepts surrounding it. Hence, by looking at the phenomena “surrounding” the action, by “expanding” our description of the action to take into account these variations, we can attain an overview of the variability of our ethical grammar regarding the action.

There are three forms this variability can take. One case is where the use of a different ethical term for an action represents *an exception* to the rule for the use of the ethical term for that action. In this case, the ethical term being used undergoes a negation (e.g. what is “wrong” becomes “right” and vice versa). For example, the act of killing someone is wrong, but killing someone in self-defense becomes “not wrong” or “justified” because the motive of self-defense conceptually overlaps with a person’s right to defend themselves against harm

and their right to live. Once again, the intermediary role played by the expression “self-defense” between the act of killing and the right to defend oneself against harm constitutes the connection between the original rule and its exception.

A second case is that of *an extension* of the rule, where the difference in the use of an ethical term reflects a difference in *degree* (in the extent of how “right” or “wrong” the action is). For example, killing someone by accident, or for revenge, or for financial reasons is wrong in each of these cases because these motives don’t conceptually overlap with any rights that would justify an exception to the original rule. Instead, what we might see in legal terminology are differentiations between degrees of murder based on such concepts as “premeditation”, “intention” etc. In such cases, while the ethical status of the action doesn’t change, these concepts account for variations surrounding the description of the act of killing someone, and in doing so, constitute the language-game that determines the use of different ethical terms for the act of killing someone, depending on whether the action was premeditated, intentional, accidental etc.

A third case is when an action has a different ethical term used for it because of disagreement within the community over the rule itself. For example, consider the debate on abortion in America during the 70s and 80s (or even in the present day). In this case, the variation in the moral/ethical status ascribed to abortion doesn’t reflect an exception to or extension of some original rule, but a social disagreement over the ethical status of abortion itself. This is due to the different language-games the speakers draw on in justifying their use of an ethical term for abortion. Hence, abortion is “wrong” in the conservative, religio-political language-game given the relation between the action and such concepts as “life” or “sanctity of life” or “state’s rights”, it is “right” in the language-game of feminism given its relation to the concepts of “bodily autonomy” or “reproductive rights”, and it is contested in the language game of civil and constitutional rights in relation to such concepts as “due process” and “right to privacy”. The inter-relationship between the concepts of “life”, “bodily autonomy”, “right to privacy” etc. defines the contours and boundaries of the abortion debate, showing that even a case of ethical disagreement isn’t completely arbitrary.

4. Shared Customs in a “Form of Life”

At the beginning, I mentioned Wittgenstein’s claim that failing to essentially define ethical concepts can make them appear arbitrary and vacuous. Oddly enough, this problem can also arise from using the method of finding transitional and intermediary links between the members of family resemblance concepts. Al Zoubi frames this “relevance” objection as the idea that “it is always possible to find similarities or resemblances between any two things”, and of the three interpretations of family resemblance he examines, it is one that he believes all of them struggle to meet (Al Zoubi 2016: 46-48). He identifies how some commentators believe that the notion “form of life” has the potential for addressing this objection by constituting relevance in terms of the shared “needs” (Bellaimy 1990) or “experiences” (Gert 1995) of a community. However, Wittgenstein’s actual answer to this worry is slightly different.

In the lecture where Wittgenstein discusses how the use of ethical or aesthetic concepts is “bound up” with the objects they modify (discussed in section 2), he goes on to state that what determines the “ideal” for such concepts is “a certain very complicated role it played in the life of the people” (AWL 1979: 36). To understand what is meant by ideal, he says, “you would need to describe the instances of the ideal in a sort of serial grouping”. Hence, the ideal “Greek profile” can be seen from the fact that “the greatest [Greek] sculptors used this form, people were taught it, Aristotle wrote on it”. That is, the ideal use of ethical or aesthetic concepts is *a social ideal observable in multiple aspects of the life of a community* (in contrast to the *philosophical* ideal that becomes the target of his criticism in the *Investigations*).

Understanding how this relates to the relevance objection involves seeing two parallels between *teaching/explaining* the ideal use of a concept and a family resemblance concept; both of which can be seen in Wittgenstein’s response to the worry about family resemblance concepts not being “bounded”:

Then it might be objected that a transition can be made from anything to anything, and so the concept isn’t bounded. To this I have to say that for the most part it isn’t in fact bounded and the way to specify it is perhaps: ‘by “knowledge” we mean these processes, and these, *and similar ones*’.

And instead of ‘and similar ones’ I might have said ‘and others akin to these in many ways’. (PG 1974: 76)

The first parallel is the *pre-established, customary* nature of the examples used in the instruction. In explaining a family resemblance concept, “[o]ne gives examples and intends them to be taken in a particular way”, and in learning it one is to “employ those examples in a particular way” (PI 2009: §71). The possibility of following this particular intention is determined by the “embeddedness” of the intention in the established customs of a community. “[A] person goes by a signpost only in so far as there is an established usage, a custom” (PI 2009: §198). To follow rules means to follow “*customs* (usages, institutions)” (PI 2009: §199), “practice[s]” with public, not private, standards of correctness (PI 2009: §202). “An intention is embedded in a setting, in human customs and institutions. If the technique of the game of chess did not exist, I could not intend to play a game of chess” (PI 2009: §337). A language requires agreement “not in opinions, but in form of life” (PI 2009: §241), not just “in definitions”, but also “in judgments” (PI 2009: §242).

The second parallel is in the teaching of the concepts *by the stating of multiple examples in a series*, and not just a *single* example. The use is constituted by similarity-relations *between* the multiplicity of examples: hence, “This *and similar things* are called...” (PI 2009: §69). The use of the concept is supported by “a whole series of props in readiness” for us to “lean on” (PI 2009: §79). Consequently, showing a similarity to any *one* of the examples isn’t enough to make the concept “vacuous”. The idea that a characteristic possessed by a single member can determine the use of a general concept is characteristic of an essentialist definition where every member possesses the defining characteristic. On the contrary, membership in a family resemblance concept is quite difficult to achieve given that the concept is constituted by the relations between a multiplicity of members. We extend the concept “as in spinning a thread we twist fibre on fibre. And the strength of the thread resides not in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres” (PI 2009: §67).

The baselessness of the relevance objection lies in a misunderstanding both these aspects. E.g. take an account by Pompa who pointed out that “street-fighting” cannot be considered a “sport” just because it is similar to

“boxing” (Pompa 1967: 66). True, but what determines this exclusion? The usage of the word “sport” that the objection is predicated on was likely taught using examples such as:

swimming, tennis, fencing, golf, weight-lifting, boxing...

When the customary usage of “sport” is made *explicit* i.e., when multiple other examples of the term are presented in a series, it becomes obvious why Pompa doesn’t see the term as applicable to “street-fighting”. This shows that those making the relevance objection against family resemblance concepts are operating under the influence of customary usages of such concepts that they have “long been familiar with” (PI 2009: §109) without realizing it. The “problem” arises from “the bewitchment” of their understanding by the fact that they use such concepts in definite ways, *despite* the absence of an essential, defining characteristic.

What does this mean for the relevant use of ethical concepts? It means that the use of an ethical term for an action is determined by the multiplicity of customs in the language-game in which the action is located within a community, and not just by any similarity the action might share with another action. E.g. We don’t call a state’s exacting a tax on its citizens “wrong” simply because it involves “taking money by force” and hence, is akin to “stealing”. The ethical status of taxation is determined by the language-game called “the constitution”, and the multiplicity of political, historical, and legal customs which structure this game. By surveying the connections between the customary use of ethical terms in various language-games within a community, one can see them as constituting the ethical framework for that community. So construed, ethical concepts are actually quite resistant to the charges of vacuity and arbitrariness directed at family resemblance concepts.

It is also important to note that the language-game in which an action is located is also customarily determined. For instance, “eating pork” or “drinking alcohol” is “wrong” in certain communities where the action is located in a religious language-game. The notion that the language-game in which an action is located is customarily determined also becomes apparent in cases of ethical disagreement *within* a community. For instance, as I mentioned in section 2, ethical disagreement over “abortion” can be traced to

different language-games such as “feminist theory”, “civil and constitutional rights”, “religio-political conservatism”, “embryology” etc. The disagreement is over which of these language-games should determine the ethical term the community uses for abortion. This is one of the reasons that I am not partial to the proposals that (on a Wittgensteinian scheme) we should classify ethical examples according to the criteria of different metaethical positions (De Mesel 2019: 140) or normative ethical theories (Kuusela 2020: 441).

References

- Al Zoubi, Odai (2016) “Wittgenstein and Family Concepts”, *Nordic Wittgenstein Review* 5, 31–54.
- Baker, Gordon; Hacker, P.M.S. (2005) *Wittgenstein, Understanding and Meaning: Volume I, Part I: Essays*. 2nd. extensively revised edition by P.M.S Hacker. Oxford: Wiley-Blackwell.
- Bellaimey, James E. (1990) “Family Resemblances and the Problem of the Under-Determination of Extension”, *Philosophical Investigations* 13, 31-43.
- De Mesel, Benjamin (2019) “Are Moral Judgments Semantically Uniform? A Wittgensteinian Approach to the Cognitivism-Non-Cognitivism Debate”, in: Benjamin De Mesel and Oskari Kuusela (eds.) *Ethics in the Wake of Wittgenstein*, New York: Routledge, 126-148.
- Gert, Heather J. (1995) “Family Resemblances and Criteria”, *Synthese* 105(2), 177-190.
- Klagge, James (2018) “Wittgenstein and von Wright on Goodness”, *Philosophical Investigations* 41, 291–303.
- Kuusela, Oskari (2020) “Wittgenstein and the Unity of Good”, *European Journal of Philosophy* 28, 428–444.
- Pompa, L. (1967) “Family Resemblance”, *The Philosophical Quarterly* 17(66), 63-69.
- Wittgenstein, Ludwig (1974) *Philosophical Grammar*. Oxford: Blackwell.
- Wittgenstein, Ludwig (1979) in: Alice Ambrose (ed.) *Wittgenstein's Lectures: Cambridge 1932–1935*. Amherst: Prometheus Books.
- Wittgenstein, Ludwig (2009) *Philosophical Investigations*, 4th. revised edition. Oxford: Wiley-Blackwell.