

JAAKKO REINIKAINEN

Making Meaning

A study in foundational semantics

JAAKKO REINIKAINEN

Making Meaning
A study in foundational semantics

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Social Sciences
of Tampere University,
for public discussion in the B1096
of the Pinni B building, Kalevantie 4, Tampere,
on 14.6.2024, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Social Sciences
Finland

<i>Responsible supervisor</i>	Professor Panu Raatikainen Tampere University Finland	
<i>Supervisor</i>	Professor Jaakko Kuorikoski University of Helsinki Finland	
<i>Pre-examiners</i>	Associate Professor Ronald Loeffler Grand Valley State University USA	Associate Professor Jeremy Wanderer University of Massachusetts Boston USA
<i>Opponent</i>	Professor Jussi Haukioja Norwegian University of Science and Technology Norway	
<i>Custos</i>	Professor Arto Laitinen Tampere University Finland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2024 author

Cover design: Roihu Inc.

ISBN 978-952-03-3420-8 (print)

ISBN 978-952-03-3421-5 (pdf)

ISSN 2489-9860 (print)

ISSN 2490-0028 (pdf)

<http://urn.fi/URN:ISBN:978-952-03-3421-5>



Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2024

ACKNOWLEDGEMENTS

As I started writing this dissertation, I operated under the assumption that Robert Brandom had mostly solved all the major issues in the philosophy of language, so all that remained was to assist him with some details and tricky objections. The point that spoke the most in favour of Brandom in my view was that he had the best aftermath to Kripkenstein's sceptical challenge, which I took to have massive importance – to be 'the' question – for concerns about meaning. By 'Kripkenstein', I mean, of course, Saul Kripke's (1982) reading of Ludwig Wittgenstein's later works.

However, in a couple of years, I came to realise, mostly due to the work of Stephen Turner (2010), that the very foundations of Brandom's project – the idea of norms implicit in practices – were not as secure as I had once thought; indeed, they had to be completely abandoned. The resulting U-turn left me (for a brief time) thinking that only radical instrumentalism conjoined with semantic anti-realism, somewhere in the region of Quine, Dennett, and Davidson, could pose a remotely plausible post-Kripkenstein framework for thinking about language, and especially about meaning.¹

Finally, likely through the subtle influence of my other supervisor Panu Raatikainen, I was led to rediscover an alternative route to realism via Kripke's (1980) theory of reference and especially Michael Devitt's associated developments, whose remark about philosophical semantics as the 'veritable Balkans of the intellectual world' (1996, 48) could not but sum up the twists and turns of my route to Sarajevo. At the same time, thanks to the fascinating work with cognitive science by my colleague Renne Pesonen (2019), I came to

¹ Personally, I blame my persistent instinct to write epic fantasy as the cause of my sweeping judgments about what is and is not possible when it comes to the curious, mythical beast we call meaning. Perhaps it is a perennial issue for philosophers at large, trying to answer too many questions at once.

reappreciate the hard virtues of Brandom's inferentialism and pragmatism, if only to simultaneously deepen my confusion with the 'normativist' aspects of the theory.

Again, reading Turner and finding his arguments (especially those targeting Brandom's transcendentalist assumptions) quite devastating made me wary of awarding 'genuine' normativity any kind of an explanatory role. For the naturalist, normativity is not a sacred cow – it is just another hamburger. The harshness of the sentiment may yet turn out a little too comfortable, but for the moment, I could not see a way around it. You cannot always choose what is best to eat, after all.

As already mentioned, very early on, I regarded the task of getting to grips with Kripkenstein's sceptical challenge as the unavoidable basic task in forming anything like general views about meaning. The problem was, of course, that the challenge itself was immensely contested and debated, its form, purpose, and origin more or less clouded in enigmas. Fortunately, there was one skeleton key to unravel the lock, which was Martin Kusch's (2006) impressive layout of the literature. Via Kusch, I could gain a glimpse of the real problems that required solving. Later, my nascent understanding of these dynamics was, if not confirmed, then supported by Jussi Haukioja's (2021, unpublished) more long-running, advanced enterprises.

Though the dissertation began as a genuine attempt to support Brandom's work, eventually I turned more critical to his approach. That being said, a major stepstone in the process for me was the opportunity to meet Brandom in person during a Viennese workshop organised on the eve of *A Spirit of Trust's* publication, April 2019. His alacrity to answer my many questions encouraged me to pursue them further.

Aside from these specific contributions, this work benefited more than words can weigh from the amicable environment of Tampere University's philosophy department and dozens of other people I met along the way. Particular gratitude goes to Aleksi Honkasalo, Jaakko Kuorikoski, Jaana Virta, Petri Räsänen, Siiri Porkkala, Paula Rauhala, Jenni Rytälä, Pasi Valtonen, Otto Sahlgren, Jani

Hakkarainen, Markku Keinänen, Jan Forsman, Tommi Vehkavaara, Arto Laitinen, Teemu Toppinen, Antti Keskinen, Corrado Piroddi, Danika Harju, Jaakko Belt, Jaan-Erik Vainio, Jaana Parviainen, Kristina Rolin, Lauri Lahikainen, Sauli Havu, Juho Rantala, and Tapio Santala.

Lastly, I wish to express my gratitude for the generous three-year funding provided by the Kone Foundation for this work.

CONTENTS

Acknowledgements	iii
List of Abbreviations	xi
Tiivistelmä.....	xii
Summary	xiv
1 Introduction.....	17
1.1 Summary of the Chapters.....	19
1.2 Chapter I	19
1.2.1 Section 2	19
1.2.2 Section 3	20
1.2.3 Section 4	20
1.2.4 Section 5	20
1.2.5 Section 6	21
1.2.6 Section 7	21
1.3 Chapter II.....	22
1.3.1 Section 9	22
1.3.2 Section 10.....	23
1.3.3 Section 11	23
1.3.4 Section 12.....	24
1.4 Chapter III.....	24
1.4.1 Section 14.....	24
1.4.2 Section 15.....	25
1.4.3 Section 16.....	25
1.4.4 Section 17	26
Chapter I: Kripkenstein's Sceptical Challenge and Its Discontents.....	27
2 Introduction to the Sceptical Challenge.....	28
2.1 Narrative by Correctness.....	30
2.2 Narrative by Justification.....	33
2.3 Comparing the Narratives.....	38
2.4 On Semantic Justification.....	44

3	Refuting Semantic Dispositionalism	50
3.1	The Problem of Finitude	51
3.2	Kusch’s Extended Defense of Kripkenstein’s Argument.....	53
3.3	The Return of Semantic Dispositionalism: cp-conditions	55
3.4	The Return of Semantic Dispositionalism: abstractions	60
3.5	Semantic Dispositionalism without cp-conditions.....	62
3.5.1	Objections to Båve.....	66
3.6	Semantic Dispositionalism without cp-conditions II	70
3.6.1	Objections to Warren	74
4	Refuting Platonism.....	77
5	Refuting Qualia	81
6	Refuting Primitivism	84
6.1	Refuting Ginsborg’s Primitivism	85
7	Lessons of the Sceptical Challenge.....	92
7.1	Kusch and the Sceptical Solution.....	94
7.1.1	Minimal Factualism	97
7.1.2	Semantic Descriptions and Prescriptions.....	99
7.2	Temporality of Meaning	103
7.2.1	On Semantic Justification Again.....	108
7.3	Causal Cures.....	110
7.3.1	Kusch’s Criticism of the Causal-Historical Response to the Sceptical Challenge	113
7.3.2	Devitt on Semantic Descriptions.....	115
7.3.3	Specifying Referential Intentions.....	118
7.3.4	Aftermath of the Challenge: Living with Indeterminacy	124
	Chapter II: All Genuine Norms Go to Transcendence	126
8	Plan for the Chapter.....	127
9	The Core Architecture.....	129
9.1	Brandom and Kripkenstein	129
9.2	Meaning as Social-Perspectival	135
9.3	Meaning as a Historical Process	139
9.4	Expressive Algorithms	146
9.5	What Are Norms Implicit in Practices?	154

10	Criticism of the Synthesis	161
10.1	The Cake Is a Lie	161
10.2	Against Normative Phenomenalism.....	168
10.3	Distinguishing between Normative Force and Content	171
11	Semantic Objectivity.....	173
11.1	Three Grades of Semantic Objectivity.....	174
11.2	A complication regarding the objectivity grades	177
11.3	Brandom’s Objectivity: Determination as a Negative Concept.....	180
11.4	External Criticism of Brandom’s Objectivity	183
11.4.1	Warrants	184
11.4.2	Truth	187
11.5	The Pragmatist Route to Semantic Objectivity of Contents Goes Via Conceptual Realism.....	189
11.6	Why Conceptual Realism Is Essential for Proper Objectivity.	192
11.7	Summary of the Section	196
12	Pragmatic Objectivity	198
12.1	On Bindingness.....	198
12.2	Three Models of Discursive Sovereignty.....	201
12.3	The Problem of Reason’s Autonomy	204
12.4	Original Error.....	210
12.5	Why Semantic Objectivity Does Not Entail Pragmatic Objectivity	212
12.6	Motivating Pragmatic Objectivity on Independent Grounds ..	216
12.7	Appearance of Necessary Reasons	225
12.7.1	Norms as Inner Reality: Response to Peregrin and Kiesselbach	229
12.8	Turner and the Disenchantment of Reason.....	239
	Chapter III: The Geist in the Machine	246
13	Introduction	247
14	Institution of Norms by Reciprocal Recognition	251
14.1	Iron, Wolves, and the Problem of Induction	253
14.2	Recognitive Hierarchy.....	256
14.3	Argument against Authority	258
14.4	Robust Recognition.....	267
15	On Experience.....	271

15.1	Four Semantic Criteria of Adequacy for Epistemic Theories ..	273
15.2	Referents as Ideal Senses	278
15.3	Comparing MIE and ST on Acknowledgement and Action....	284
16	On Recollection	288
16.1	Introducing Sittlichkeit and Modernity	291
16.2	The Problem of Alienation: An Initial Look.....	296
16.3	Overcoming Alienation I: Introducing the <i>Kammerdiener</i>	302
16.4	Overcoming Alienation II: Four Meta-Meta-Attitudes	305
16.5	An Interlude Summary: Alienation and the Problem of Finitude 309	
16.6	Overcoming Alienation III: The <i>Geist</i> in the Machine	314
16.6.1	Reading 1 of the Always Already argument.....	317
16.6.2	Reading 2 of the Always Already Argument.....	321
17	Back to Kripkenstein	327
17.1	The Causal-Historical Account and the Four Semantic Criteria for Epistemic Theories	335
18	Conclusions	339
	References	341

List of Tables

Table 1.	Models of discursive sovereignty.....	268
Table 2.	Four meta-meta-attitudes.....	306

LIST OF ABBREVIATIONS

- WRPL — Kripke, S. (1982). Wittgenstein on rules and private language: An elementary exposition. Harvard University Press.
- PI — Wittgenstein, L. (1953). Philosophical investigations (Anscombe, G.E.M., Trans.). Oxford.
- PP — Brandom, R. (2011/PP). Perspectives on Pragmatism: Classical, Recent, and Contemporary. Cambridge MA, Harvard University Press.
- ST — Brandom, R. (2019a). A spirit of trust: A reading of Hegel's phenomenology. Harvard University Press.
- TMD — Brandom, R. (2002). Tales of the mighty dead: Historical essays in the metaphysics of intentionality. Harvard University Press.
- AR — Brandom, R. (2001). Articulating Reasons: An Introduction to Inferentialism. Cambridge MA, Harvard University Press.
- MIE — Brandom, R. (1994). Making it explicit: reasoning, representing, and discursive commitment. Harvard University Press.
- BSD — Brandom, R. (2008). Between saying and doing. Oxford University Press.
- RP — Brandom, R. (2009). Reason in philosophy: animating ideas. Harvard University Press.

TIIVISTELMÄ

Tämä on työ kielen filosofian alaan lukeutuvassa metasemantiikassa. Sen tarkoituksena on auttaa vastaamaan kysymykseen siitä, miten sanat saavat merkityksensä. Työ koostuu kahdesta osasta. Ensimmäisen osan tarkoitus on puolustaa väitettä, että lukuisista yrityksistä huolimatta nk. Kripkensteinin skeptiseen haasteeseen, erityisesti ns. äärellisyyden ongelmaan, ei toistaiseksi ole annettu toimivaa suoraa ratkaisua. Toisen osan tarkoituksena on kriittisesti tarkastella Robert Brandomin filosofiaa, jota tulkintani mukaan voidaan tarkastella vastauksena skeptiseen haasteeseen. Väitän, että vaikka Brandomin nk. normativistinen lähestymistapa tarjoaa periaatteellisen ratkaisun äärellisyyden ongelmaan, kohtaa se joukon muita ongelmia, joiden valossa se on hylättävä.

Mikä on Kripkensteinin skeptinen haaste? Nimi 'Kripkenstein' on kirjallisuudessa käytetty lyhennös Saul Kripken Ludwig Wittgensteinin myöhempiä töitä tulkitsevalle pitkälle esseelle, joka julkaistiin 1982. Esseessään Kripke esittää (Wittgensteinin esittävän) merkityskeptisen haasteen, joka vaatii selittämään, miten millään sanalla voi olla määrätty merkitys. Haasteen katsotaan yleisesti koostuvan kolmesta ongelmasta, jotka ovat: (1) äärellisyyden ongelma, (2) virheen ongelma, ja (3) normatiivisuuden ongelma. Martin Kuschia seuraten katson äärellisyyden ongelman olevan näistä keskeisin. Äärellisyyden ongelman voi kuvailla lyhyesti seuraavasti. Kaikki luonnollisen kielen puhujat ovat äärellisiä olentoja: heillä on vain äärelliset taipumukset käyttää kielellisiä symboleja kuten sanoja ja matemaattisia symboleja kuten '+'-merkkiä. Jokaisen aktuaalisen kielenkäyttäjän jokaisen ilmaisun kohdalla voidaan siis kysyä, käyttääkö hän tiettyä merkkiä säännön S_1 vai säännön S_2 mukaan. Esimerkiksi '+'-merkin yhteydessä voidaan kysyä, käyttääkö hän sitä yhteenlaskun vai 'yhteenlaskun' säännön mukaan. Yhteenlasku antaa kahden luvun summaksi saman tuloksen kuin yhteenlasku paitsi tietyn mielivaltaisen raja-arvon ylittävillä luvuilla, jota hypoteesin mukaan yksikään subjekti ei ole laskenut tai tule koskaan laskemaan. Jokaisen kielenkäyttäjän aktuaalinen matemaattinen toiminta on hypoteesin

mukaan empiirisesti yhteensopivaa sekä yhteenlaskun että qyhteenlaskun kanssa. Koska jokaisen subjektin kohdalla tulee vastaan myös tämän matemaattisten taipumusten kohdalla raja, jonka ylittäviä lukuja tämä ei ole laskenut tai tule koskaan aktuaalisesti laskemaan, on skeptikon mukaan epämääräistä, tarkoittaako kukaan aktuaalisesti koskaan '+'-merkillä yhteenlaskua vai qyhteenlaskua.

Tulkintani mukaan äärellisyyden ongelmaa ei kirjallisuudessa ole Kripken ja Kuschin työstä huolimatta ole otettu niin vakavasti kuin pitäisi. Tällä taas on merkittäviä seurauksia yleiselle teorialle metasemantiikassa koskien sanojen merkitysten alkuperää ja luonnetta. Erityisesti esitän, että metasemantiikan tulee käsittää merkitykset ajallisesti määräytyneinä. Lisäksi väitän, että Kripken kehittämä kausaalis-historiallinen viittaamisen teoria voi toimia epäsuorana vastauksena skeptiseen haasteeseen, sekä selittää, kuinka ainakin jotkin merkitykset voivat olla ajallisesti määräytyneitä.

Brandomin filosofian eräs keskeinen pyrkimys on antaa vastaus äärellisyyden ongelmaan tarjoamalla ratkaisu normatiivisuuden ongelmaan. Väitän kuitenkin, että Brandomin versio kirjallisuudessa yleisesti käytetystä 'merkityksen normatiivisuuden' hypoteesista ei voi toimia siten kuin hän haluaa. Keskeiset ongelmat ovat monisyisiä, mutta lyhyesti ilmaistuna Brandomin haasteena on antaa periaatteellinen vastaus versioon Agrippan trilemmasta, eli kuinka jatkuvien oikeutusten ketju voi päättyä johonkin muuhun kuin (i) jo käytettyihin premisseihin, (ii) primitiiviseen premissiin, tai (iii) äärettömään premissien ketjuun. Tulkintani mukaan Brandomin vastaus trilemmaan ei ole kelvollinen.

SUMMARY

This is a work in the philosophy of language and metasemantics. Its purpose is to help answer the question about how words acquire their meanings. The work is divided into two parts. The purpose of Part One is to defend the claim that, despite numerous attempts, the so-called Kripkenstein's sceptical challenge, and especially the problem of finitude, has not been offered a successful straight solution. The purpose of Part Two is to critically examine Robert Brandom's philosophy, which can be treated as an answer to the sceptical challenge in my interpretation. My main claim is that although Brandom's so-called normativist approach does provide a principled solution to the sceptical challenge, the proposal faces a host of other problems in the light of which I reject it.

What is Kripkenstein's sceptical challenge? The name 'Kripkenstein' is an abbreviation used in the literature for Saul Kripke's reading of Ludwig Wittgenstein's later works, published in 1982. In the long essay, Kripke presents (that Wittgenstein presented) a sceptical challenge that demands an explanation of how any word can have a determinate meaning. The challenge is usually thought to consist of three interrelated problems: (1) the problem of finitude, (2) the problem of error, and (3) the problem of normativity. Following Martin Kusch, I view the problem of finitude as the most important one. According to the problem, briefly, since any actual subject is a finite being, i.e. capable of exhibiting only finite expressions of linguistic signs, it will always be possible to ask whether, in a given context, she is following a rule S_1 or rule S_2 that corresponds to the meaning of the sign. For example, supposing that the sign in question is '+', it can be asked whether the subject in a given context has followed the addition rule or an alternative quaddition rule, according to which any addition problem exceeding certain arbitrary cardinality results in five. According to the hypothesis, the actual use of '+' by any speaker is compatible with both rules. Since any speaker has only finite dispositions to the mathematical conduct, there will always be some limit after which there is no fact of the matter, according

to the sceptic, which rule the subject will have turned out to have followed. Hence, it is indeterminate whether any actual subject has ever calculated according to the addition rule or the quaddition rule.

From my view, despite the work of Kripke and Kusch, the problem of finitude has not been taken as seriously as it ought to have been. This in turn has significant consequences for the metasemantic theory concerning the origin and nature of meaning. In particular, I claim that the way how the meaning of words is determined must be understood as temporal in nature. Moreover, I claim that Kripke's causal-historical account of reference can work as a non-straight solution to the challenge that explains how at least the meanings of some words can be temporally determined.

One central aim of Brandom's philosophy is to give a straight answer to the problem of finitude by providing an answer to the problem of normativity. However, I claim that Brandom's version of the 'normativity of meaning' hypothesis cannot work as he intends. The key issues are difficult to summarise, but to put it succinctly, the crucial problem concerns Brandom's ambition to give an answer to a version of Agrippa's trilemma, or how a chain of justifications can end in anything else than (i) premises that have already been used, (ii) a primitive premise, or an (iii) infinitude of premises. My claim is that Brandom's resolution to the trilemma is not sound.

1 INTRODUCTION

This is a work in foundational semantics, nowadays often called metasemantics. Its orienting question is: 'Why does a word have the meaning that it in fact has?'

In more particular terms, the study is divided into two parts. Part One concerns Kripkenstein's sceptical challenge, its interpretation and aftermath. Part Two concerns Robert Brandom's (MIE; ST) philosophy of language, its interpretation and relation to the sceptical challenge. Below, I shall elaborate on these parts and their relation.

Through this work, the most important idea that I want to convey to the reader is that there is no straight solution to the so-called *problem of finitude* presented by Kripkenstein's sceptical challenge. On this point, I remain convinced that the original arguments – as presented by Kripke, reinforced by Kusch (2006), and extended by myself – have not been taken as seriously as they ought to have been in the massive literature that has grown around the sceptical challenge. There are several reasons for why the problem of finitude has been comparatively neglected in the literature, which are discussed at the beginning of Chapter I.

But what is the problem of finitude? Let me give the unfamiliar reader a brief anticipatory exposition. Assume that Jones, who has been learning to do addition, is now facing a novel problem of 'How much is $58 + 67$?'. Furthermore, assume that there is a bizarre sceptic who challenges us to explain in virtue of what fact Jones has – until now – been following the addition function as opposed to the so-called 'quaddition' function, according to which any answer to an addition problem exceeding 57 is five. What the problem of finitude amounts to is the observation that, given that Jones is a finite being with finite arithmetic dispositions, he cannot calculate with numbers exceeding certain cardinality. He simply is not disposed to perform calculations with *very* large numbers. However, since the limit of 57 is arbitrary and could be moved to the region of very large numbers, it follows that

Jones' arithmetic dispositions cannot determine whether, in a given context, he is indeed following the addition function or some other, seemingly arbitrary function.

Rejecting the possibility of a straight solution to the problem of finitude obviously leaves us with the question of what comes next. One important entailment of the rejection advanced in this work is that meanings, facts grounding meaning facts, and meaning ascriptions must be finitely determined. They must be temporally determined. What this means is that the semantic values of expressions (e.g. the addition function for '+', or the set of all tables for 'table') cannot be fixed by logical rules that would delineate partitions for every logically possible application of the expression such that it would be determinate for every application whether it was correct or incorrect. As we shall see in Chapter I, Section 7, this rejection of *absolute determinacy* entails the rejection of the idea that the meaning of a declarative sentence would be given by its truth conditions. Furthermore, in the same section, I attempt to offer a replacement picture for truth conditions based on the causal-historical account of reference that is also compatible with the *historicity of meaning*.

What about the two other problems, of which the sceptical challenge is usually taken to consist of, namely that of error and especially normativity? While I believe that the problem of error is at bottom an aspect of the problem of finitude, it is with the problem of normativity that we connect to the two chapters discussing Brandom's work. In short, Brandom, who invests great hopes for understanding language as the idea that it is fundamentally a kind of *genuinely normative practice*, can be read as someone attempting to give a straight solution to the problem of finitude by providing an answer to the problem of normativity. While 'genuine normativity' is a tricky notion to pin down, the idea is that meanings must be such that they can give *good reasons* for using the expression one way rather than another, or that facts grounding meaning facts must be such that they *justify* the use of the expression, or that ascriptions of meaning must be such that they (purport to) attribute *responsibility* and *authority* over various meaning sentences (e.g. '+' means addition) to the speaker. For what it is worth, I think these are good ways to cash out the vexing issues concerning normativity inherent in the sceptical challenge. However, I ultimately aim to reject the idea of genuinely semantic or discursive normativity from theoretical consideration, though strictly speaking, the critical main thesis concerns only Brandom's version of semantic normativism. That being said, many points of

controversy are mutually shared by other authors in this vein, if only because Brandom has influenced them.

1.1 Summary of the Chapters

Having presented my main theses above, I shall now mechanically summarise the arguments for them, as they will resurface in detail. As a general note, despite the impression of definitiveness the reader may have gotten until now, I do not take any of my conclusions to be completely secure in their reasons. There is simply too much literature, which I either could not discuss here, or which I simply am ignorant of to justify a claim to completeness. Yet, I remain convinced enough to undertake the responsibility for defending the conclusions against further objections, rather than displaying them as speculative conjectures in no need of such defence.

Since I did not wish to write the work twice over, these summarised descriptions will likely appear sketchy and cryptic, especially to anyone who is not already very familiar with the topics. Nonetheless, I think they will be helpful in order to see the big picture, at least after reading the entire work. (Note that here, I omitted the various introductory and summary sections.)

1.2 Chapter I

1.2.1 Section 2

In this section, I introduce Kripkenstein's sceptical challenge first in its 'raw form', i.e. the form in which Kripke casts it, followed by what I consider the two most prominent narrative interpretations of the challenge. I call these *narrative by correctness* and *narrative by justification*. A few different points distinguish these narratives about the argumentative form of the sceptical challenge, but the vital issue is which of the three major problems (the problems of normativity, error, and finitude) is the most important one. As I understand them, narrative by correctness emphasises the role of the normativity problem in the challenge, especially as Boghossian's (1989) sense of 'correctness conditions', whereas narrative by justification, mainly on account of

Kusch (2006), takes the problem of finitude to be the crucial pivot. I will argue why the justification narrative is the stronger exegetical reading of the sceptical challenge. Moreover, I argue that despite the numerous attempts that will be examined in the following four sections, the sceptical challenge remains unsolved.

1.2.2 Section 3

In what is perhaps the most important section of this chapter, I will defend the sceptical arguments against different solutions, specifically addressing the problem of finitude. (Briefly, a solution counts as straight if it defends a certain philosophical picture, or a major portion of the picture, argued in Section 2 to be the implicit *target* of the challenge.) I continue to rely heavily on Kusch's arguments, mostly extending and defending them against later objections. The principal opponent is called *semantic dispositionalism*, arguably the most important straight candidate for solving the problem of finitude. The major claim of this section is that semantic dispositionalism, even in its most recent iterations, remains unable to properly solve the problem of finitude and the challenge.

1.2.3 Section 4

In this section, I examine the strategy called *Platonism* that gives a straight solution to the problem of finitude. The section is short because I will argue that the problems of Platonism are like those of semantic dispositionalism. These two strategies can thus be undermined together, as Kripke effectively does.

1.2.4 Section 5

In this section, I examine *the qualia strategy* that gives a straight solution to the problem of finitude. Although *the qualia strategy* is quite different from the previous strategies, I believe we can dismiss it without delving into it extensively in this work.

1.2.5 Section 6

Section 6 will conclude the discussion of the four strategies that attempt to give straight solutions to the problem of finitude. The fourth strategy, called *primitivism*, is in a sense the most confounding of the bunch, and would require a lengthy discussion to be fully appreciated. However, for reasons that I will not discuss in this summary, primitivism does not provide a straight solution to the problem of finitude. I shall also briefly discuss Hannah Ginsborg's (2011; 2021) recent strategy for dealing with the sceptical challenge that has some primitivist tendencies, deeming her case interesting but ultimately rejecting it.

1.2.6 Section 7

The main conclusion drawn from the previous sections is that there is no straight solution to the problem of finitude. This conclusion is incomplete, for I have not discussed nearly all arguments and objections that either have been or could be made against my own. However, I believe the arguments are potent enough to draw the conclusion, and to take it seriously. The purpose of the seventh section is to understand what that entails.

To begin with, one important entailment of the claim that there is no straight solution to the problem of finitude is that the ascriptions of meaning sentences (i.e. sentences of the rough form 'S means x by 'y') are not meaningful in virtue of expressing propositions with truth conditions. At the beginning of the section, following Kusch, I argue that this does not mean that the sceptic trips into a kind of self-contradiction, where they would deny that any word has any determinate meaning. The crucial point is that, although the ascriptions of meaning sentences do not have 'robust' truth values based on truth conditions, they can still be considered as 'minimally truth-apt', which avoids drawing the self-contradictory conclusion that no word would have any determinate meaning.

Some authors sympathetic to the sceptical arguments, Kusch among them, have thought that if the ascriptions of meaning sentences can have no truth conditions, it will be pointless to talk about 'meanings' in the sense in which analytic philosophy has for several decades now, i.e. a sense in which meaning would correspond to a

property of certain linguistic items. David Bloor's (1997) concept of 'meaning finitism' is a famous example of this type of thinking. However, this section aims to argue that there is still room to hold a realist view on meaning, even within the framework of meaning finitism, which we are forced to consider due to the conclusion that there is no straightforward solution to the problem of finitude. Contrary to Kusch, I will argue that this way can be found by recourse to the causal-historical theory of reference, inaugurated by Kripke and developed further by Michael Devitt (1996; 1999) and Mario Gómez-Torrente (2019). That being said, it will be necessary to adjust that theory, or picture, in view of the sceptical challenge's lessons. How exactly that is to be done will largely be left for future work to figure out, though I will give a few stabs at it at the end of Chapter III.

The idea that meaning could be identified as a property of certain linguistic items faces independent challenges from semantic expressivists, who claim, roughly, that the ascriptions of meaning sentences are expressive rather than descriptive speech acts. My reason for focusing on Brandom's brand of semantic expressivism is that his account is, in my view, deeply motivated by the sceptical challenge, particularly by the problem of normativity. Indeed, one of Brandom's main slogans is that meaning corresponds, not to a kind of property, but to a kind of propriety of linguistic items. In a way, Brandom's goal is to give a straight solution to the problem of finitude by solving the problem of normativity, of which he has an understanding much like the justification narrative. Elaborating and criticising this normativist strategy is the purpose of the next two chapters.

1.3 Chapter II

1.3.1 Section 9

This section introduces Brandom's systematic thinking in the philosophy of language, most importantly in MIE, in the context of the sceptical challenge discussed in Chapter I. In focus is the question how compatible Brandom's *discursive scorekeeping* account is with the problem of finitude. As it will turn out, these are indeed compatible with each other, for Brandom's work can be read as an extended,

non-straight solution to the sceptical challenge that takes semantic normativity as its keystone.

A key theme of my reading of Brandom is that he attempts to construct a grand synthesis between what will be known as *phenomenalism about norms* and *normative phenomenalism*. To put it briefly, the first commitment states that all (discursive) norms are products of our activities, of our takings and treatings of norms as being in force. The second commitment balances this reductive claim by stating that our treatings and takings are *genuinely responsible* to the norms themselves, considered as somehow transcending our practical grasp on them. The norms are said to be *instituted* by the practices, which then *confer* propositional contents on our expressions and linguistic items.

1.3.2 Section 10

This section examines the criticisms of Brandom's account, with a focus on the synthesising theme mentioned earlier. I think Brandom's critics can be roughly divided into two camps: those who think that normative phenomenalism should prevail over phenomenalism about norms and vice versa. This division is certainly not exhaustive, for Brandom's topics are many as are the objections directed against him, but it is a useful orienting distinction to make in the context of the problem of objectivity towards which the chapter moves.

1.3.3 Section 11

Regarding the problem of objectivity that Brandom's synthesising project faces, in this section, I shall first address its semantic side, in contrast to the pragmatic. The idea is that in contrary to many of his critics, Brandom is able, already in MIE, to offer a *principled* defence of semantic objectivity, or how meaningful expressions can be said to *represent* objects beyond the discursive practice. To explain this is to explain a lot in favour of the synthesising project, or how our practical takings of norms to be in force can transcend the practice in the sense that we become genuinely responsible to the norms themselves, and by extension to objects in the world. The major argument here is that the missing piece to make semantic objectivity work as

intended is *conceptual realism*, which amounts to the idea that the world itself is conceptually structured. Although Brandom does not offer an independent defence of conceptual realism in MIE, I argue that if true, it suffices to solve the problem of semantic objectivity as he lays it out for himself.

1.3.4 Section 12

The pragmatic side of the problem of objectivity, which forms the key issue for Brandom's synthesising project, concerns *genuine normative force* as opposed to genuine representation. Whereas the previous four sections are principally expositional in nature, the chapter's original argumentation brings its weight to bear here, for I shall argue that Brandom is indeed unable to solve the problem of pragmatic objectivity in MIE. To address the *problem of reason's sovereignty*, we need more than just a genuine representational relation in scorekeeping. For the *contents* of the norms implicit in the discursive practice could be genuinely representational without their *force* being genuinely obligatory or authoritative, I argue.

MIE is not the only place where Brandom has sought to solve the problem of reason's sovereignty. The last chapter of the work takes on the formidable *A Spirit of Trust*, which represents both a complimentary advance and a radically new direction in relation to MIE from my standpoint.

1.4 Chapter III

1.4.1 Section 14

In this section, I shall look at the formal side of Brandom's new solution to the problem of reason's sovereignty. At the heart of the solution is *the Hegelian model of reciprocal recognition* that, in my reading, tries to synthesise the earlier models of *Kantian Autonomy* and *Queen's Shilling*, discussed in the previous chapter. These models offer different answers to the question of how discursive norms can be said to be implicitly instituted by practices. Although the new solution is fundamentally continuous with the ideas and terminology presented in MIE, it is also significantly different, for

besides a formal solution, Brandom offers a genealogical account of how the community capable of exhibiting implicit normativity could have originated from more primitive natural capacities, a question notoriously left unanswered in MIE. I believe that the genealogical account has challenges similar to MIE regarding the development of normative abilities from non-normative ones.

1.4.2 Section 15

In this section, I examine Brandom's interpretation of Hegel's concept of *experience* (*Erfahrung*) that plays an important role in ST. Once we have offered a 'formal' solution to the problem of reason's sovereignty in the previous section, what remains is a substantial question concerning the *symmetry* of reciprocal recognition. In short, the formal solution to the problem of reason's sovereignty is that if recognition is *de jure* transitive and *de facto* symmetrical, A's authority to recognise B's authority is made authoritative by B's reciprocal recognition of A's authority. This may appear circular, but only if we are not told how recognition can be *de facto* symmetrical, which requires answering how it can be temporally extended. For (discursive) authority to be genuinely normative, it must be recognised. Since recognitions of authority must be authorised to be genuine, the way to avoid blatant circularity is that some recognitions become authoritative *retroactively*.

In this section, I will also briefly compare MIE and ST regarding normative attitudes and their explanatory roles.

1.4.3 Section 16

In this section, I will look at Brandom's interpretation of Hegel's *recollection* (*Erinnerung*), which plays a major role in the account along with experience. Put shortly, recollection is supposed to explain how reciprocal recognition can be *de facto* symmetrical by explaining how recognition can be not only *retrospective*, but also *retroactive*. This is the most important section of the chapter, its main substance being the *Always Already argument*, as I call it. Brandom's argument for how recollection can have retroactive normative effects and make reciprocal recognition symmetrical is that as scorekeepers, we are necessarily committed to *forgiving* each other's discursive

infringements. To forgive in this sense means to recognise the authority of others in applying concepts in judgements.

In my view, the Always Already argument, at least in the forms in which I am able to construe it, is unsound and perhaps invalid. A sceptic who refused to recognise discursive norms as genuinely binding would not be committing a *pragmatic contradiction*, which the Always Already argument essentially demands is the case. They do not, by what they say, contradict what they do while saying it.

1.4.4 Section 17

In the final section of the chapter, I return the discussion to Kripkenstein's sceptical challenge and my own proposed, non-straight solution to it that draws from the causal-historical account of reference. The idea here is to compare my non-normative account with Brandom's normativist one, but not only in a negative spirit. For I believe that Brandom's account in ST has something important to offer to the causal-historical account, even if the causal-historical account must reject his ambitious metaphysical justification narrative culminating in the Always Already argument. The main claim here is that there is no such thing as genuine semantic normativity, not at least in the sense that Brandom fathoms, and neither is there a theoretical need to explain genuine semantic normativity in the sense of the justification narrative discussed in Chapter I. This result, too, is incomplete, yet strong enough to merit the assertion, I believe.

CHAPTER I: KRIPKENSTEIN'S SCEPTICAL CHALLENGE AND ITS DISCONTENTS

2 INTRODUCTION TO THE SCEPTICAL CHALLENGE

Imagine there is a subject named Jones, who has been learning under instructions to do addition. Jones quickly grasps adding small numbers, and with some effort, tackles multi-digit ones too, occasionally making errors. Currently, he has been tasked to add '58' and '67' together. Everyone, including Jones himself, expects him to give the correct sum, which is 125, when a bizarre sceptic suddenly enters the scene and asks:

'Why should Jones answer "125" to "58 + 67" and not "5"? What makes this answer correct in the context of the training he has received so far? What is the fact justifying the judgement that instead of learning the addition function, Jones has in fact been taught to follow a completely different function all along – let's call it "quaddition" – which returns any output to addition problems exceeding the input 57 with the "quum" that is 5 instead of the sum?'

This is roughly how Saul Kripke (WRPL) begins his sceptical challenge of meaning gathered from a reading of Ludwig Wittgenstein's later works, most importantly (PI). For four decades, the challenge has fuelled the debate in the philosophy of language, and – for all intents and purposes – remains unsolved. The main purpose of this chapter is to examine reasons for why that is, and how the fact might be changed.

One reason that makes the challenge so thorny is that there has not been a broad consensus in the literature about what the argumentative structure of the challenge truly is, or whether there even is a single coherent structure. Although most commentators have recognised that the challenge involves three major problems – finitude, error, and normativity – it has been much less clear how exactly these problems relate to each other. Confusion about the argumentative structure has led to disagreement about how the major problems, especially that of normativity, actually should be formulated.

There are at least two major competing narratives in the literature about how the argumentative structure of WRPL should be understood, or what the 'shape' of the challenge is. Furthermore, I believe the key difference lies in how we approach the normativity problem, which is widely recognised as being hard to explain. In effect, one group of commentators takes normativity primarily as a problem of *justification*, while others take it as a problem of *correctness conditions*, as I shall explain below.

This chapter conveys my stance that the correctness narrative is exegetically wrong about WRPL in key respects, while the justification narrative, especially as advanced by Martin Kusch (2006), is essentially right on the key exegetical points. Once we have the shape of the challenge clearly in view, it becomes possible to evaluate both the sceptical arguments and objections to them with more definiteness. In Sections 3, 4, 5, and 6, I shall argue against what will be specified in Section 2.2 as the four main strategies of responding to the sceptical challenge. What these strategies share is that they seek to answer the sceptic by formulating the truth conditions of meaning sentences, such as 'S means addition by '+' and then showing why they are (if they are) fulfilled with Jones.

Finally, in Section 7, I shall focus on the lessons of the challenge under the assumption that it has not, and indeed cannot, be refuted by the truth-conditional strategies. I believe there are three basic avenues of proceeding from there. The first is given an approximate outline by Kripke himself in his reconstruction of the sceptical solution. The spirit of the sceptical solution is to downgrade our ambitions regarding the extent to which (foundational) semantic questions can be systematically pursued in favour of promoting attention toward general pragmatics instead; essentially, how we use expressions, such as 'means that', as opposed to what we mean by them.

The second option seeks to preserve the possibility of systematic semantic theorising by radically revising our understanding of what 'means' means. In effect, some authors, such as Robert Brandom, Allan Gibbard, and Huw Price, think that the root mistake is to think that the predicate 'means that' would have the pragmatic function of attributing properties, or that meaning sentences could be analysed as *descriptions*. Instead, meaning is taken to play an *expressive, pragmatic* function, where 'pragmatic

function' means explaining what it is that we *do* in using semantic vocabulary, like the locution 'means that'.

The third option, one that I shall seek to develop further, adopts a kind of middle ground between these two positions. Against the sceptical recoil from systematic semantics altogether, a way forward will be sketched to think (foundational) semantic questions in a non-deflationist spirit. Against the expressivists, it will be argued that meaning sentences and the 'means' predicate can be analysed as doing descriptive work, although not in the sense presumed by the truth-conditional framework. In particular, I believe we should give up the idea that the descriptivity of sentences (i.e. the reason they count as 'descriptions') rests on expressing propositions with truth conditions.

An alternative picture, if not a theory, will be sketched along the lines of the causal theory of reference inaugurated by Kripke (1980). Following certain developments of Kripke's founding ideas, I claim we can think of at least some meanings as (a) non-descriptive, (b) empirically accessible in a broad sense, and (c) ontologically constituted by the causal-historical chains of reference.

The focus of my account along these lines will be to show that it is compatible with the sceptical challenge and, to some extent, with the sceptical solution as well. The purpose is not to deliver a complete theory in a brief space, but to build towards a better picture than what is left of the truth-conditional strategy once the sceptical challenge has taken its course.

2.1 Narrative by Correctness

The narrative that focuses on correctness conditions starts with Paul Boghossian's (1989) interpretation, which has a straightforward understanding of the normativity problem:

Suppose the expression 'green' means *green*. It follows immediately that the expression 'green' applies *correctly* only to *these* things (the green ones) and not to *those* (the non-greens). The fact that the expression means something implies, that is, a whole set of *normative* truths about my behaviour with that expression: namely, that my use of it is correct in application to certain objects and not in application to others [...] The

normativity of meaning turns out to be, in other words, simply a new name for the familiar fact that, regardless of whether one thinks of meaning in truth-theoretic or assertion-theoretic terms, meaningful expressions possess conditions of *correct use*. (On the one construct, correctness consists in *true* use, on the other, in *warranted* use.) (Boghossian 1989, 513)

What leads Boghossian to understand semantic normativity in terms of correctness conditions? One reason is that he thinks that according to Kripkenstein², normativity plays an important 'criterion of adequacy' role for foundational semantic theories. One aspect of the correctness narrative is to take Kripke to be laying down such a criterion, which must be met by any successful foundationalist semantic theory. Alternatively, the correctness narrative may interpret the normativity problem as a metaphysical thesis concerning the nature of meaning. I take it that nothing too crucial hangs on which form the challenge is taken: the major point is that the correctness narrative takes 'semantic normativity' to be something which WRPL actively advances as a necessary *explanandum* for (foundationalist) semantic theory.

The correctness narrative further splits into two subplots that are commonly, and somewhat loosely, known as 'normativist' and 'anti-normativist' respectively. A key point of contest between normativists and anti-normativists is whether correctness conditions, as depicted by Boghossian (1989), involve prescriptive entailments in some typically unspecified semantic vein. 'Prescriptive entailments' is understood in terms of normative force canonically denoted in English by expressions, such as 'ought', 'may', 'is obliged to' or 'is prohibited to'. Boghossian is a great exemplar of this debate because he has defended both sides of it at different times: in his work (1989), the connection between 'correctness' and 'ought' is taken to be near-synonymous (see especially: 1989, 509, 530), whereas later (2003; 2005), he changed his mind and became critical of the easy association between 'ought' and 'correctness'.

² 'Kripkenstein' is a conventional name in the literature to abbreviate 'Kripke's reading of Wittgenstein's Philosophical Investigations'. It is important to be aware that the views and arguments attributed to Kripkenstein do not (necessarily) belong either to Kripke or Wittgenstein, as Kripke cautions the reader at the beginning of the book (WRPL, 5). My use of the name 'Kripkenstein', or sometimes the term 'the sceptic', is in part meant to emphasise a non-committal stance on exegetical issues regarding the origin of the sceptical arguments.

An important watershed in this debate is Anandi Hattiangadi's paper (2006) in which Hattiangadi makes several key moves and distinctions that have become staples in the literature. Among these is the regimentation of some of the central 'normative formulations' in the preceding literature as biconditionals, starting from Boghossian's (1989) idea of correctness conditions as the key sense in which meaning is normative:

$$\textit{Meaning Platititude (MP): } t \text{ means } F \rightarrow (x)(t \text{ applies correctly to } x \leftrightarrow x \text{ is } f)$$

In this equation, ' t ' means a term, ' F ' its meaning, and ' f ' whatever feature (or collection thereof) in virtue of which F applies to x . According to Hattiangadi, MP is accepted nearly by everyone save for certain sceptics, among them Quine and Kripkenstein himself (2006, 222).

The two further key biconditional regimentations are:

$$\textit{Correctness: } S \text{ means } F \text{ by } t \rightarrow (x)(S \text{ applies t 'correctly' to } x \leftrightarrow x \text{ is } f).$$

and

$$\textit{Prescriptivity: } S \text{ means } F \text{ by } t \rightarrow (x)(S \text{ ought (to apply } t \text{ to } x \leftrightarrow x \text{ is } f)).$$

Hattiangadi's main claim is that *Correctness* does not entail *Prescriptivity* or any of its emendations, and hence that semantic normativity is not a problem for the broadly naturalist foundationalist theories of meaning (2006, 225). However, 'semantic correctness' is to be interpreted; it does not by itself entail any of the problematic prescriptive truths or oughts, which she takes Kripkenstein to argue to be implied by the ascriptions of meaningful states. She follows Boghossian's original interpretation in that what *Correctness* picks up is merely the old foundationalist question – in virtue of what do our words have the meaning ('correctness conditions') they actually do – but denies that Kripkenstein would show some special normative underpinnings of this question that would implicate a novel 'semantic ought'. Hattiangadi's argumentative strategy closely resembles Hume's law to the effect that 'is' cannot entail 'ought', or that one cannot derive prescriptive consequences from merely descriptive premises. Since *Correctness* is taken to be merely descriptive of semantic facts, it alone cannot be the source of prescriptive

norms. Yet, if *Correctness* is assumedly essential to meaning due to MP, then meaning is not normative in the sense of *Prescriptivity* or its variations without further premises being involved. A similar approach has also been used earlier by Glüer and Pagin (1999), who frame their argument in terms of the Searlean (and Rawlsian) distinction between regulative and constitutive rules, and later by Glüer and Wikforss (2009).

The response by normativists has been varied, and I have examined some of them in my work (2020). One important further distinction due to Glüer and Wikforss (2009), that has been made in the correctness narrative, concerns the explanatory order of meaning and prescriptive norms. Is it that semantic facts entail prescriptive consequences as with *Prescriptivity*, or is it that semantic facts themselves are determined by norms?

What is noteworthy is that the farther branches of the correctness narrative seem to have little substantial connection to the original arguments presented in WRPL. In a way, the correctness narrative has developed its own gravitation where the arguments for and against semantic normativity can be considered having independent interest regardless of their links to Kripkenstein's sceptical challenge. If that is indeed the case, I do not think either the challenge or the correctness narrative are worse off for that. However, my interest here is concentrated on uncovering what I take to be the strongest exegetical interpretation of WRPL, which I believe is outlined better by the justification narrative.

2.2 Narrative by Justification

The other major narrative about the argumentative structure of WRPL centres on the idea that *justification* is the key form that normativity takes, not correctness conditions. But what exactly is supposed to be justified, by (and to) whom, how, and when?

The case for the justification narrative is made most powerfully by Martin Kusch (2006). According to his reading, the sceptic's primary question concerns the justification of what I shall tentatively call 'ordinary applications'. Simply put, ordinary applications mean the everyday practice of using words meaningfully, e.g.

when talking about the weather or what to have for dinner. This encompassing category also covers our arithmetical practice of adding. The first question of the sceptic is what justifies Jones' answer of '125' to the addition problem '58 + 67' (WRPL, 8-9). Where our *ordinary practice of arithmetical justification* is concerned, the case could not be clearer: Jones has given the correct answer because 125 is the sum of 58 and 67 and because the numerals '58', '67', and '125' have their ordinary denotations.

Now, it is crucial to observe that the sceptic's main challenge does *not initially* appear at the ordinary level of application. In short, the sceptic does not doubt whether 125 really is the sum of 58 and 67 in the 'ordinary arithmetical sense'. Instead, what the sceptic first challenges is whether Jones' use of the sign-type '+' in the past has 'tracked' the addition function as opposed to the alternative hypothesis of the quaddition function. This is what Kripke calls the 'metalinguistic sense' of justification. This level is best understood as the question of what justifies, not our ordinary arithmetic applications, but our *ascriptions* of ordinary applications in the form of meaning sentences, such as 'Jones means *addition* by '+'.

Before we continue, let us clarify the connection between regular application and the metalinguistic level. To be sure, there is nothing 'out of the ordinary' in the metalinguistic level in the sense that we frequently ask each other questions of the form 'What did you mean by that?' or 'What does this word mean?' Indeed, I think it is important to realise that the metalinguistic level can be understood as already implicated in the ordinary level of justification, where we use ordinary reasons to answer the aforementioned questions (e.g. clarifying my intentions or looking up the word in a dictionary). Again, the sceptic does not doubt the very possibility of checking a word's meaning from a dictionary. Rather, the challenge is supposed to target our *philosophical* or *metaphysical* understanding of meaning and semantic justification, not the ordinary folk practices, much like Hume's scepticism of induction and causation was not meant to stop us from trusting in our ordinary causal judgements about the world (WRPL, 63). In other words, the sceptical challenge is first and foremost a challenge for a *theory* of meaning, not something which the ordinary folk practice of meaning ascriptions would have to deal with.

What, then, is the relation between the ordinary and the metalinguistic levels in the 'philosophical' sense? In short, at stake is recognisably the foundationalist line of questioning: in virtue of what do our words have the meanings they actually do? What does meaning consist of? How do we know what we mean by our words? Are there any facts that determine what meaning our words have? Whereas the ordinary metalinguistic issues of justifying meaning sentences, e.g. by consulting a dictionary to deal with local, parochial problems, the philosophical metalinguistic questions demand more encompassing answers. But why should such abstract issues apply to the original target of the sceptic, namely the specification of Jones' mathematical behaviour? Why would not the ordinary metalinguistic methods and reasons, e.g. asking Jones whether he meant to add or to quadd by using '+', be sufficient in delivering a justification for meaning sentences describing him? After all, sceptical conundrums aside, the ordinary practice strongly seems to work out well most of the time.

To rush ahead slightly, I believe one of the key lessons of the sceptical solution is to refute a certain view about the relations of philosophical metalinguistic (or metasemantic) theorising about ordinary practices and the ordinary practices themselves. What the sceptical solution effectively questions is both the *need* and *possibility* for a *certain way* of theorising that seeks to somehow 'rise above' the ordinary practices. One way to understand this 'rising above' would be to showcase a philosophical foundationalist justification for ordinary applications, i.e. a justification that itself was in principle independent of ordinary applications. Of course, in a trivial sense, any philosophical story to that effect depends on at least *taking* the ordinary justifications for word applications to be *presently* good, as is established by the 'ground rules' of the challenge (WRPL, 11-12). In particular, the sceptical challenge brings to doubt the possibility of giving any kind of truth-conditional account of meaning sentences. However, if the very possibility of explaining the justification of ordinary ascriptions of meaning sentences by truth-conditional means turns out to fail, that should soon enough make us conclude that there can be no reasonable need to offer such an account, for otherwise we would be led to the 'insane and intolerable' (WRPL, 60) conclusion that no word has no determinate meaning. (To clarify, an ascription of a meaning sentence is justified in this sense when one has justified that it has determinate truth conditions.)

Now, we may take a broad overview of the shape of the sceptical challenge seen by the justification narrative. The sceptic wants to know what justifies our ordinary applications (including our ordinary scientific and arithmetical applications of terms). This is the primary target of doubt. The primary *way* to doubt the primary target is by asking what justifies the metalinguistic level of applications, i.e. the practice of ascribing determinate meanings to each other's utterances by meaning sentences of the form 'S means x by 'y''. If the ascription of meaning sentences *cannot* be justified, the thought goes: neither can the practices of ordinary justifications. For then it becomes impossible to justify any answer to a simple addition problem because one cannot justify any metalinguistic convictions needed to formulate a determinate answer: Jones' answer '5' is equally justified as '125' might be. Observe, however, that even if the sceptical challenge cannot be answered, all we lose is the philosophical or metaphysical route to justifying ordinary applications. As we saw above, the ordinary level already includes the ordinary methods of justifying the ascriptions of meaning sentences, which the sceptic nowhere challenges this directly. It is only if we assume that the ordinary level *requires* a philosophical/metaphysical justification in order to *be* justified, and not just be universally *taken* as justified, that we are led to the incredible (and self-defeating) result that no word has no determinate meaning.

What are the main philosophical strategies of answering the sceptical challenge for showing non-ordinary justifications for ascribing meaning sentences? Insofar as WRPL is concerned, they are:

- 1.) The dispositionalist strategy
- 2.) The Platonist strategy
- 3.) The qualia strategy
- 4.) The primitivist strategy

Here, I shall briefly illustrate the four strategies before discussing them individually in the oncoming sections. The main result will be that all of them face severe difficulties in trying to justify our ordinary practice of ascribing meaning sentences to each other *insofar as the ascriptions are understood as expressing propositions with truth conditions*.

The most important strategy, and probably the most popular, is the dispositionalist strategy. A natural candidate for the truth conditions is the semantic dispositions of subjects, which can be described in mental or non-mental terms. The dispositionalist strategy is arguably also the strongest contender in the league, which is why Section 3 will be the longest as it discusses this very strategy.

In contrast, as in the original discussion of WRPL, the Platonist strategy shall receive comparatively little attention here. The reason, to be elaborated in Section 4, is that the semantic dispositions of subjects are arguably a necessary part of any truth-conditional account. If it can be shown that semantic dispositionalism fails, I believe it inevitably follows that Platonism does as well.

Together, dispositionalism and Platonism represent the two major contenders in what I think can usefully be called the *externalist* camp of justification. A defining trait of 'externalism' as I use the term is a disregard on the condition that Jones himself should have epistemic access to the facts or reasons justifying his answer to the mathematical question posed to him for the justification to be genuine. Instead, both Platonism and dispositionalism think that truth alone can do most or all of the work for justifying the ascriptions of meaning sentences to Jones in the sense that the ascriptions can be justified even if no one has epistemic access to the facts or reasons, which serve as truth-makers for the ascriptions. So long as we can get a firm grasp of what it means for meaning sentences to be true, the externalist can settle, e.g. for reliabilism, to explain why we are justified in our ordinary practice of ascribing meaning sentences to each other and consequently in our ordinary applications.

The contrasting 'internalist' strategy of justification is best represented by certain 'old-school' empiricist doctrines that attempt to justify meaning sentences by appealing to reasons that are (a) mentally accessible to the subject doing the justification and (b) not dependent on dispositionalist truth conditions, which include modal terms traditionally considered problematic by empiricists. Instead of truth conditions described in modal externalist language, the internalist appeals to internal or proximal states of the subject, e.g. her phenomenal or observational states, to deliver the justifications. As with Platonism, this strategy does not get much attention in WRPL as well, mostly because it appears unappealing at the outset and

because it has been powerfully criticised on independent grounds elsewhere, most famously by W.V.O. Quine (1951) and Wilfrid Sellars (1956).

Finally, we have the primitivist strategy, arguably the most confusing of the bunch, as signalled by Kripke's remark that '[s]uch a move may in a sense be irrefutable, and if it is taken in an appropriate way Wittgenstein may even accept it' (WRPL, 51). In the same breath, Kripke also calls primitivism 'desperate' since it states that the truth conditions of meaning sentences are *sui generis* and unanalysable by philosophical, scientific, or ordinary means. In Section 6.0, I shall argue that primitivism is not an appealing position.

2.3 Comparing the Narratives

In this section, my aim is first to summarise how the two narratives described above differ from each other. Second, I argue that the justification narrative is a better exegetical interpretation of WRPL. Third, I shall entertain some reasons for explaining why the narratives have diverged as they have.

As I already mentioned, the key difference between the narratives concerns the place and nature of the normativity problem in the argumentative structure of the challenge. Briefly, the correctness narrative involves Kripkenstein actively pursuing an argument about the essential normativity of meaning. In contrast, the justification narrative sees him advancing a critical *reductio ad absurdum* argument, rejecting the four main philosophical strategies one by one. These strategies aim to resolve the sceptic's primary question concerning the justification of ordinary applications of words through the philosophical justification of meaning sentences. In sum, the source of the narrative difference is whether normativity is understood as a 'criterion of adequacy' for the foundationalist semantic theory in the sense that 'any proposed candidate for being the property in virtue of which an expression has meaning must be such as to ground the normativity of meaning—it ought to be possible to read off from any alleged meaning-constituting property of a word, what is the correct use of that word' (Boghossian 1989, 509).

According to this 'normativist' line of thinking, 'semantic correctness' is understood as immediately entailing certain prescriptions (oughts or mays) that are in force for the speaker by virtue of her meaning something by a word. Furthermore, at stake is a special 'semantic prescription' that is comparable yet distinct from moral, epistemic, prudential, etc. prescriptions. The assumption that explaining the existence of such semantic normativity is the criterion of adequacy imposed on the theories of meaning by WRPL forms the core of the correctness narrative and both of its subplots. The main difference between the subplots is whether they consider WRPL to succeed or not in this effort. Their common assumption is that Kripkenstein never gives up the idea that for a word to have meaning, some specifically semantic norms must be in force for the speaker; the *reductio* argument against, e.g. semantic dispositionalism, works *on the back of* the normativity requirement.

Precisely, the opposite is true according to the justification narrative. 'Semantic normativity' does not pose a criterion of adequacy for foundational semantic theories because semantic normativity is part of the *target* of the main *reductio* argument.

The exegetical reasoning for supporting the justification narrative on this point was first appreciated by José Zalabardo (1997). What he emphasises is that *if* Kripke advances the *reductio* argument against semantic dispositionalism *by* appealing to semantic normativity, something like the following argument would have to be involved:

- (A) The fact that determines the satisfaction conditions of a predicate licenses evaluative claims about ascriptions of the predicate.
- (B) Dispositional facts are descriptive.
- (C) Descriptive facts cannot license evaluative claims.
- (D) Therefore dispositional facts cannot determine the satisfaction conditions of predicates. (Zalabardo 1997, 469)

Premise (A) roughly corresponds to the MP condition of Hattiangadi (2006) discussed above. Along with (B), it can be safely considered being trivially true, and

together with (C), the argument is valid.³ In effect, all the weight for the soundness of the argument rests on premise (C), which is exactly where the rub of Zalabardo's exegetical argument kicks in, for nowhere in WRPL does Kripke explicitly argue for the 'Humean injunction' (C) (1997, 478). However, (C) is by no means trivial, and especially not so with specifically semantic evaluations or prescriptions. So, either we are to conclude that Kripke simply missed the justifications for the key premise of what allegedly is his most powerful (reconstruction of the) argument against semantic dispositionalism, or then the argument from (A) to (D) is not Kripkenstein's.

Further support for Zalabardo's reading comes in his recognition that Kripkenstein's major *reductio* against semantic dispositionalism (or the broader truth-conditional strategy) comes in the form of the extensional infinitude problem (or problem of finitude), on which I shall focus in Section 3. For now, however, it suffices to note that the glaring absence of justifications for (C) in WRPL forms a powerful argument against the exegetical cogency of the correctness narrative.

Kusch also argues against what I have here called the correctness narrative reading about WRPL. He agrees broadly with Zalabardo's alternative interpretation, which puts justification in the centre of the normativity problem as opposed to correctness conditions, but Kusch goes beyond Zalabardo by claiming 'justification' to cover a much richer notion in WRPL than the internalist variant, which Zalabardo discusses. The richer notion includes internalist, externalist, and primitivist forms of justification corresponding roughly to the four strategies, as mentioned in the previous section. I shall have more to say about this richer notion of justification below. Kusch's main arguments, which I will not repeat here, against Boghossian's correctness narrative are that (a) this narrative ascribes to Kripkenstein's weak arguments and is hence uncharitable, and (b) there is no explicit argumentative support for the correctness narrative in WRPL (Kusch 2006, 51-64).

Although I take Kusch's arguments to be cogent, one might be left with an impression that they leave it mysterious that anyone could have ascribed the

³ In fact, (B) is not trivial, e.g., to an expressivist about modal and dispositional terms, but the complication can be ignored here since in the context of the challenge, dispositional terms are taken as descriptive.

correctness narrative to WRPL to begin with. The worry then becomes that by depicting the justification narrative as too obvious, Kusch's criticism itself turns out to be uncharitable to its target. Therefore, the failure to make Boghossian's (1989) misreading understandable would count as a point against its status as a misreading.

Luckily, I believe there is a simple explanation for where Boghossian's reading goes wrong. The reason that leads Boghossian to the correctness narrative is that he undervalues the crucial *temporal* dimension of the challenge, which is emphasised, e.g. in the early readings of Colin McGinn (1984) and Crispin Wright (1984). For example, McGinn writes that:

The notion of normativeness Kripke wants captured is a transtemporal notion: it is the idea of present use being in accord with past meaning; a linguistic mistake accordingly consists in no longer using a word with the meaning originally intended (without of course intending to introduce a change of meaning by explicit stipulation). We have an account of this normativeness when we have two things: (a) an account of what it is to mean something at a given time and (b) an account of what it is to mean the *same* thing at different times – since (Kripkean) normativeness is a matter of meaning now what one meant earlier. (1984, 174)

Boghossian criticises McGinn's reading for failing to explain the status of the normativity problem as a criterion of adequacy on foundational semantic theories, for according to him, any theory, and especially semantic dispositionalism, could trivially satisfy it 'since there are perfectly determinate facts about what dispositions are associated with a given expression at a given time' (Boghossian 1989, 513). Against Wright's idea that correctly applying an expression depends on its past applications, comparatively to how correctly castling in chess depends on what previous moves have been made, Boghossian argues that Wright's reading makes meaningful application depend on mental rule-following, which already presumes that mental expressions have correctness conditions because 'rule-following' in the relevant sense for Boghossian is already intentional, meaningful activity (Boghossian 1989, 517).

What Boghossian essentially claims is that any notion of *sameness of meaning* across temporally distinct applications of a word trivially falls out from the true answer to what he, along with Simon Blackburn (1984, 281-2), considers being the primal issue, namely in virtue of what is an application of a word semantically correct or incorrect. The thought is natural enough, for it surely seems that we can intelligibly ask and

answer to what Jones means *now* independently from what he has meant in previous 'nows' by the same word. What matters more for Boghossian is the correctness of *application*, which is a relation between a word and its referent, and not the sameness of *use*, which he assumes can be understood on the basis of the application relation (cf. Kusch 2006, 61). This is analogous to thinking that the diachronic issue of the sameness of meaning across time depends logically on the synchronic issue of correct application, which, again, is a perfectly natural way to approach the problem. I shall return to the temporal dimension of the sceptical challenge, which I take to be crucial to solving it, in Section 7.

However, the synchronic-first approach is not how Kripkenstein poses the challenge. Recall that the primary target of doubt is how Jones ought to answer the addition problem *presently* posed to him. The sceptical challenge does not go on by doubting the ordinary justification for his present answer, but by doubting, in the metalinguistic sense, the meaning that *has* previously governed Jones' applications of '+'. In other words, the question 'How Jones ought to go on presently and in the future with the '+-sign?' is seen to depend in part not only on his current intention but also his *past* uses of '+', how he has actually used it *before*. Not all the past uses necessarily count as correct applications, but the point is that the correctness of the applications can only be determined once we know what meaning actually has been individuated by his use of '+', or which arithmetical function he has actually been tracking. This is how rule-following, or 'how to go on the same way', is essentially connected to the problem of meaning understood as the determination of correctness conditions.

But why is it wrong to consider the past use simply as comprising the individual, synchronic applications, all of which can in principle be considered independently from each other as to their meaning *qua* correctness? The reason is that Kripkenstein provides an argument precisely against the idea, which Boghossian (1989) takes to be trivial, that there would be determinate facts about how anyone is disposed at any given moment to apply any term. In other words, Boghossian appears to think that Kripke's main claim is that the relation between semantic dispositions and correctness conditions is indeterminate, whereas in my view, the key argument is that semantic dispositions themselves are indeterminate. That is the heart of the problem of finitude, on which I shall focus in Section 3. Indeed, I think that part of the reason

Boghossian does not take the temporal elements in Kripke's exposition seriously is that he undervalues the objections against dispositionalist theories via the problem of finitude. Instead of taking Kripke's arguments at face value here, Boghossian opts for advancing 'a sweeping argument' of his own against dispositionalism, which centres on the prescriptive reading of 'semantic correctness' (1989, 537).⁴

To sum it up, the main reason that I think leads Boghossian (1989) to the correctness narrative is that he overlooks the temporal dimension of the challenge, and the reason he overlooks it is that he does not take Kripkenstein's arguments against the problem of finitude seriously enough, whereas Kusch does. But what the problem of finitude precisely shows is that it is ultimately wrong to consider the meaning of a word as something comprising correctness conditions of applications, which could be in principle examined synchronically and without relation to one another. This conclusion, which can be affirmed only once we have appreciated the true depth of the problem of finitude in Section 3, will be returned to in Section 7.

To finish this section, I wish to consider one more reason that might have influenced Boghossian's choice of narrative. The root reason can be considered being not just argumentative, but also expositional. In his presentation of the challenge, Kripke retained much of the dialectical structure of *Philosophical Investigations*, in which several avenues of answering the sceptic are considered in brief intervals. In one particularly popular paragraph, Kripke writes that:

The dispositionalist gives a descriptive account of this relation: if '+' meant addition, then I will answer' 125'. But this is not the proper account of the relation, which is *normative*, not descriptive. [...] The relation of meaning and intention to future action is *normative*, not *descriptive*. (WRPL, 37)

Boghossian (1989, 530) is by far not the only one who has quoted this section to support the correctness narrative. Indeed, taken out of context, that seems to be exactly what Kripke is saying, i.e. implicating the claim (C) as presented by Zalabardo. But as already noted, Kripke never actually delivers *justifications* for this claim, as he merely states it. Moreover, in the very next paragraph, he continues by reminding the reader that the original problem concerns the *justification* of the

⁴ As a matter of fact, Boghossian (2012, 44) has later admitted that he originally underestimated the force of the problem of finitude.

ascription of meaning sentences, such as 'Jones means addition by '+''. Judging by the context, 'justification' is meant in the internalist sense, where Jones himself is supposed to have mental access to the reasons in order for the justification to be legitimate: but answering that 'I am disposed to answer '125'' hardly provides legitimate (internalist) justification of the answer because one usually has no mental access to one's arithmetical dispositions. This is the sense of 'normative' that Kripke is arguably after in the oft-quoted section. This observation also leads Zalabardo to his alternative reading, which is correct, although not the whole truth about justification as Kusch argues.

2.4 On Semantic Justification

The previous section sought to portray the shape of Kripkenstein's sceptical challenge by contrasting two narratives about its general argumentative structure. The major conclusion argued for is that the shape should be understood in terms of the justification narrative as opposed to the correctness narrative. In short, the argumentative structure of WRPL is to construct a *reductio ad absurdum* against a collection of ideas that Kusch calls 'Low-brow Meaning Determinism', or MD_{LB} (2006, 4). In Kusch's interpretation, corresponding to every rejected thesis of MD_{LB}, WRPL offers an alternative idea, which together make up the 'sceptical solution' that Kripke offers in Chapter 3 of WRPL (Kusch 2006, 16). I shall avoid a thorough discussion of MD_{LB} because most of the theses are irrelevant to the problem of finitude that is my main topic of interest in the oncoming sections.

My goal in this section is to elaborate on the concept of justification, as it appears in WRPL, in order to further specify the justification narrative. Although I agree overtly with Kusch's reading, my version will have a somewhat distinct emphasis. Namely, whereas Kusch delivers a more even-handed treatment of MD_{LB}, I shall focus on semantic justification as the backbone of MD_{LB} and thus the backbone of the challenge itself. The ensuing discussion is meant not so much to compete with Kusch's reading as to complement it. In Section 7, where the temporal dimension of the challenge will come into focus, I shall further complement Kusch's reading by identifying a certain core structure in the MD_{LB} picture that explains why at least some of its component theses go together.

I think that a useful way to approach semantic justification is by first laying out all the different senses of justification that I think are present in WRPL, and then resuming to see how they relate to each other. Once we have the overview in sight, I shall proceed by supporting the argumentation of WRPL regarding concerns about justification in Sections 3 through 6.

The first salient distinction concerning justification in WRPL is about *the object* of justification, or what is being justified. In Section 2.2, I already showed that the primary targets of sceptical doubt are the ordinary applications of words or expressions, e.g. our arithmetical practices involving the symbol '+' that we commonly use to denote the addition function. In order to justify the ordinary applications, the sceptic claims, we must be able to justify the ascriptions of meaning sentences that (purport to) describe what we are doing in ordinarily applying words. Therefore, the secondary objects of sceptical doubt are the ascriptions of meaning sentences and their justification. The justification of meaning sentences can also be called justification in the metalinguistic sense as opposed to the justification of first-level applications.

Related to the distinction between the metalinguistic and first-level justification is that between the *ordinary* and *philosophical* justification mentioned in Section 2.2. In order to elaborate this distinction, we have to return to the ground rules of the challenge. In Kripke's exposition, rather than using the terms 'ordinary' and 'philosophical' justification, the ground rules are formulated in terms of 'past' and 'present' applications in the sense that the sceptic does not *initially* challenge our present use of '+' to denote the addition function. This is meant to 'avoid confusing questions about whether the discussion is taking place 'both inside and outside language' in some illegitimate sense' (WRPL, 12).

In my view, Kripke's mode of exposition that frames the ground rules of the challenge in temporal terms serves *both* as a shallow expositional device *and* cues us to a deeper problem. It is a shallow device because ultimately, the sceptical problems are shown to concern our present applications as well, which become past as soon as they are finished. Yet, the distinction between past and present applications does not simply vanish in the challenge's course because unless some words presently keep their usual meanings, we cannot even raise the sceptical challenge. Clearly

enough, unless the words 'What fact determines that Jones means addition rather than quaddition by '+'?' mean what they ordinarily mean, or at least are taken to mean what they ordinarily mean, the challenge cannot be expressed or thought about *at any time*.

As I already mentioned, we have ordinary means to answer metalinguistic questions about the meaning of words, ranging from using a dictionary to more painstaking anthropological efforts to translate unfamiliar expressions like 'gavagai'. These methods are 'ordinary' not because they assume (by necessity) that some words have their ordinary meanings, but because they do not rest on an explicit theory that could justify the methods used in a foundationalist sense. In contrast, we turn to philosophy and metaphysics to find such a theoretical justification for why we are indeed justified, at least for the most part, to take our words to have their ordinary meanings. Kusch argues that sceptical doubt targets the very possibility of formulating a sufficient philosophical theory along these lines, but *not* the validity of ordinary applications, which forms the centre of the sceptical solution's inherent primitivism. The primitivism is incompatible with any kind of explicit theory about foundational questions in semantics, yet does not entail the arguably self-refuting conclusion that no word has the meaning it is ordinarily taken to have, as I shall explain in Section 7. There, I shall also argue why the sceptical solution is unsatisfactory and how the philosophy of language can do better than that.

The object of justification includes semantic descriptions and prescriptions. We recall that the first question that the sceptic presents is why Jones *ought* to answer '125' instead of '5' when queried the answer to ' $68 + 57$ '. The ought is clearly a prescription, but of what kind, precisely? Is it merely an instrumental ought that follows from the fact that the symbols Jones applies have their ordinary meanings, and that Jones' intention is to calculate according to the addition function? Or is a more profound kind of ought at stake, one that ensues solely from the meaning of the symbols without mediation by Jones' intentions and other attitudes? What would count as a justification of the prescription mandating the response '125' clearly hinges on the 'ought' in question.

For now, we can sideline the issues regarding the nature of 'ought' and settle on contrasting it with the justification of the descriptions that the sceptic asks for

because of one answer to prescriptive justification. Perhaps the most natural response is to say that the 'ought' is instrumental in nature, following simply from the usual meanings of the symbols Jones uses and from his intention to add. In other words, what in part justifies the (instrumental) prescription is a certain description of Jones' intentions and his behaviour with the '+' sign, namely the ascribed meaning sentence 'Jones means *addition* by “+”.' That is a semantic (untensed) description of Jones' behaviour with the '+', which does not yet obviously entail any semantic prescriptions about what Jones ought to do with '+'.

Now, a third salient distinction about justification enters the picture: the temporal register of justification. For what the prescriptive issue of justification naturally targets is how Jones ought to behave with the '+' sign *in the future*, considering the hypothesis, ascribed to him by the meaning sentence 'Jones means *addition* by “+”', that he has been adding *in the past* and together with his *present* intentions. Here, it is good to recall that, according to the correctness narrative, the temporal distinctions are in principle irrelevant to the question of what it would be semantically correct for Jones to do. Once we know what it is for Jones' untensed (or atemporal) application of '+' to be correct or incorrect, i.e. what determines his meaning of addition as opposed to quaddition by '+', we can generalise the answer to all possible temporal moments and see in which of them he means *addition*, in which something else.

There are two more classes of distinction regarding justification that are worth noting. Fourth, justification can be *internalist* or *externalist* in kind.⁵ Fifth, it may matter who performs the justification, i.e. is it Jones himself while he ponders, in the first-person mode of deliberation, what answer he ought to give to the addition problem presented to him? This is the sense in which Jones is *guided* in his behaviour by the meaning of the '+' sign. In contrast, Jones' second or third-personal audience can be said to *assess* his behaviour in light of the meaning they take the '+' sign to have.

Putting all this together, we get the following taxonomy of justification:

⁵ I have used (and continue to do so) the terms 'internalism' and 'externalism' in a rather cruder fashion than in which they are usually encountered in epistemological debates, trusting that this simplification is more helpful than harmful. Although I think this seminal distinction is implicit in Kripke's text, his grain of analysis is far from displaying all the subtleties that a fuller discussion would have to endure.

1. The object of justification, or *what* is being justified.
 - a.) Either certain *descriptions* or then certain *prescriptions* regarding Jones' applications of '+'.
 - b) Either the *first-order applications* of '+' or then *ascriptions* of ordinary applications of '+' in meaning sentences, or the metalinguistic sense.
2. The temporal register of justification, or *when* the object of justification takes place. Both meaning sentences and first-level applications can be read in *tensed* (temporal) and *untensed* (atemporal) senses. The correctness narrative reads them in the untensed, atemporal sense.
3. The nature of justification, or *how* it proceeds.
 - a) Semantic justification can be *externalist* or *internalist* in kind.
 - b) Semantic justification can be *ordinary* or *philosophical/metaphysical*.
4. The subject of justification or *who* delivers it. Semantic justification can be performed either by (first-person) *guidance* or then by (second or third-person) *assessment*.

These classes can be cross-combined in various ways. For example, Jones himself may appeal either to externalist or internalist reasons to justify either his first-level applications of '+' or then his self-ascriptions of meaning sentences involving '+', and either as descriptions of his past applications or as predictions of his future use. He may himself serve as the assessor of his past applications, too.

How are these distinctions related to each other in the argumentative structure of WRPL? From the previous discussion, we recall that, for the correctness narrative, an important question is whether (and in what conditions) semantic descriptions entail semantic prescriptions in a non-instrumental sense. Normativists answer yes, whereas anti-normativists answer no. I have argued that whatever merits that debate has, they are not directly relevant to the arguments of the sceptical challenge, which revolve around a distinct set of problems.

The key problem for the challenge proper is whether and how the ascriptions of meaning sentences can be justified as *descriptions*. MD_{LB} understands semantic descriptions as propositions with truth conditions, which then leads to the four

major strategies of justifying meaning sentences brought up in Section 2.3. This task forms the backbone of the challenge, of which I am most interested in this chapter.

My main line of argument in Sections 3 through 6 seeks to show that MD_{LB} 's understanding of meaning sentences as descriptions expressing propositions with truth conditions is wrong. It is wrong because the ascriptions of meaning sentences cannot be justified as descriptions with truth conditions: all four main MD_{LB} strategies of truth-conditional justification are refuted successfully in WRPL. What I shall deliver here, with the help of Kusch, extends these arguments in response to certain objections that have been made over the years. Because of the volume of literature, the discussion will not be complete, but I think the case made is powerful enough to warrant drawing the aforementioned conclusion. The implications of MD_{LB} 's failure as an account of the justification of the ascription of meaning sentences is picked up in Section 7.

3 REFUTING SEMANTIC DISPOSITIONALISM

Above in Section 2.2, I stated that the most important, and probably the most popular, respondent to the sceptical challenge is known in the literature as 'semantic dispositionalism'. The core idea of dispositionalism is that the truth conditions of meaning sentences can be identified with the dispositions of subjects to apply words in certain ways and not others. The reason semantic dispositionalism has been so popular is taken to be the following: The sceptic wants to know what justifies the ascription of meaning sentences, part of which is to explain what makes a given meaning sentence true. In order to know whether the sentence 'Jones means addition by '+' is justified, we need to know (a) the conditions in which the sentence would be true and (b) whether the conditions obtain in the actual world. Semantic dispositionalism is the most important respondent of the sceptical challenge because it is a natural way to understand the truth conditions of meaning sentences, and because truth appears as the simplest way to justify the ascription of meaning sentences. Surely enough, it is to be expected that whatever truth there is to be had about Jones' applications of '+', at least some of these truths must be *about* Jones in some sense. Furthermore, it is insufficient to appeal as evidence only to how he actually applies the '+' sign because (a) Jones might make a mistake in the calculation and (b) because of his finite nature, there will always be some addition problems that he has never considered and therefore might reveal that, in fact, he is consistently disposed to quadd and not add.

While semantic dispositionalism has gained popularity, I agree with Kusch that it is unable to resolve the problems it confronts.

Parts of the early version of this chapter were published in Finnish in my work (2021b).

3.1 The Problem of Finitude

First, it will be helpful to see how Kripke originally presented the problem of finitude (WRPL, 26-28). The sceptic wants to know what facts make it so that Jones has meant in the past, means at the present, and will mean in the future addition by '+' and not quaddition in the sense that he will produce sums and not quums. The dispositionalist first claims that there is an intrinsic state of Jones, picked out either by a mental or non-mental disposition, that will cause him to give the sum to any addition problem where his intention regarding the use of '+' remains the same. Kripke disagrees with the dispositionalist's claim that a simple conditional disposition can solve any addition problem, pointing out that finite beings cannot add extremely large numbers. An immediate remedy is suggested: *ceteris paribus*⁶, given any addition problem, Jones *would* give its sum and not quum where he means addition by '+'. The *ceteris paribus* condition is a necessary addition to the dispositionalist claim that Jones adds precisely because literally no one can be expected to give the sum of any addition problem, roughly in the same way we cannot expect any actual gases to behave according to the letter of the Kinetic Theory of Gases. Certain unrealistic abstractions or idealisations (more on those later) must be made, e.g. to the properties of the molecules, such as their elasticity and spatiality. However, the Kinetic Theory of Gasses has proven itself to be scientifically respectable, which is a powerful argument for its truth. Perhaps something similar applies to Jones' dispositions to add?

Kripke saw two problems with this answer. First, it is not clear what it would *actually* mean to idealise Jones' dispositions so that he could add enormous numbers. Where extra brain stuffing, magical elixirs to prolong life indefinitely, and unlimited working memory are needed, the dispositionalist explanation seems to succumb to science fiction. We cannot know what would actually happen if such conditions took place, whether Jones would actually give the sum or go insane and offer the quum instead. Hence, positing 'extra brain stuffing' onto the left-hand side of the *ceteris paribus* conditional disposition cannot be shown to entail the sum of any addition problem on the right-hand side. It does not help to claim that our ignorance about the actual

⁶ Literally, "all things being equal". The basic idea of *ceteris paribus* conditions is that a certain causal effect will follow a certain causal input in the absence of certain defeating conditions. For example, if I strike this match, it will catch a spark unless it is wet.

effects of adding extra brain matter and prolonging Jones' life indefinitely concerns only whether he would, in those conditions, remain (or become) an adder or a quadder, for what is challenged is the very notion that 'enhanced Jones' would behave in *any* determinate way with the '+' sign after the intervention and not simply go insane, for instance.

The second problem occurs when the dispositionalist changes from a conditional disposition to a counterfactual one:

If I somehow were to be given the means to carry out my intentions with respect to numbers that presently are too long for me to add (or to grasp), and if I were to carry out these intentions, then if queried about ' $m+n$ ' for some big m and n , I would respond with their sum (and not with their quum). (WRPL, 28)

Here, it is left unspecified just what the means would amount to, i.e. whether it is a magical elixir or unlimited working memory that is needed and how the required intervention is to be actualised. Nonetheless, a failure ensues because the counterfactual formulation makes the dispositionalist answer trivial by being circular. In response to the counterfactual dispositional formulation, the sceptic can ask what the intention is to be carried out with the unspecified means. The answer cannot be *assumed* to be 'the intention to add' because that makes the counterfactual circular. All things being equal, the sceptic has the equal right to assume that in fact, the intention is to quadd⁷ because all the finite evidence we may survey underdetermines the function actually at play. Here, the claim that Jones himself knows *a priori* what his intentions are is also brought into question, for the sceptic can simply repeat his challenge: what facts determine that Jones has such *a priori knowledge* about *addition* and not *quaddition*? After all, a finite being cannot possibly have 'mentally checked' all the possible answers he would give to any addition problem beforehand. Likewise, it is insufficient to specify the intention as 'going on the same way as before', for the set of Jones' previous behaviour with the '+' sign must then be specified in non-intentional terms for the proposal to evade circularity.

In sum, the problem of finitude leads the dispositionalist to a dilemma where the required *ceteris paribus* clause is empirically unjustified or then turns the proposal circular. The dispositionalist can either rely on unrealistic but logically possible

⁷ I prefer this over 'skadd(idition)', for uniformity.

extensions of Jones' dispositions or use a counterfactual interpretation of the disposition. In the first case, the *ceteris paribus* clause is empirically unjustified because we do not actually know if there is any determinate way in which Jones would behave were the unrealistic interventions realised, while in the second case, the clause is unjustified because it leads to a circular specification of Jones' intentions regarding the '+' sign.

3.2 Kusch's Extended Defense of Kripkenstein's Argument

Kripke's exposition, which was just reviewed, has drawn a host of objections. In this section, my purpose is to examine some of these in an outline while also showing Kusch's arguments against them. The central theme here concerns the status of *ceteris paribus* clauses and whether they really cannot support semantic dispositionalism. I think Kusch does mostly excellent work in applying the sceptical thinking in showing that they cannot.

One famous attempt to rebut the sceptic rejection of *ceteris paribus* clauses comes from Jerry Fodor (1991). Fodor's argument is analogical in form: it seeks to undermine the sceptic by claiming that if he is right, much of the current science that employs *ceteris paribus* laws (e.g. Kinetic Theory of Gasses) will have to be rejected since they, too, make unrealistic assumptions in their conditional and counterfactual claims. Since it is (at least nomically) impossible for molecules as described by the Ideal Gas Law to be perfectly elastic or to occupy zero space, the law is essentially counterfactual. Nonetheless, sceptical considerations aside, we have every empirical reason to believe that it is a true counterfactual. Hence, there must be something wrong in the sceptic's rejection of *ceteris paribus* conditions with dispositions to mean something by an expression because otherwise, the same arguments would undermine most of our science.

Kusch's strategy is to disqualify the analogy by offering three dissimilarities between *ceteris paribus* laws (henceforth 'cp-laws') as they are used in the sciences and what the dispositionalist tries to make of them in the case of meaning. The dissimilarities (named by me) are:

Systematicity. Scientific cp-laws belong to systems of interrelated laws, but it is unclear to what if any system of established laws the dispositionalist proposal should be classified; it is *ad hoc* to claim the status of scientific laws to the disposition to add enormous numbers.

Approximation. Although the scientific counterfactual laws rely on strictly speaking false assumptions about the world, the data values predicted by the laws can be experimentally approximated to better fit the ideal; but it is unclear how the disposition to add with enormous numbers could be approximated, by experiment of otherwise, to our actual behaviour. 'Someone who sees an approximation here strikes me as someone who thinks that my leaping into the water from a one-metre-high diving platform approximates my 'leaping' into the centre of the universe.'

De-idealisation. Scientific cp-laws can be de-idealised to better correspond to actual data and to make more accurate predictions but it is unclear how to de-idealise the dispositions to add enormous numbers to better match the data. (Kusch 2006, 102)

There is also a fourth, more complicated dissimilarity that Kusch considers that is presented as a modification of Fodor's position. Most laws in special sciences have cp-conditions because reality's complexities tend to disrupt regularities beyond the fundamental level of physics. An independent problem that this creates is to find a justified demarcation between three types of failures for any cp-law: (a) flawed research setting, (b) random interference from the world not exactly understood by the experimenter, and (c) genuine exceptions to the law. The ability to distinguish between (b) and (c) is seminal, for only (c) obliges a revision or rejection of the law. Kusch argues that the dispositionalist law-candidates cannot distinguish between (b) and (c) on principled grounds; thus, they are not genuine cp-laws. The reason is that unlike many genuine cp-laws, the dispositionalist proposal is 'cognitively permeable', i.e. it is sensitive to the background beliefs of Jones, which can vary indefinitely. Jones' disposition to utter 'horse' in the presence of horses can misfire not only on a misty field (i.e. because of external interferences), but also if he believes that the cows around here look like horses, or many other strange beliefs, as pointed out by Boghossian (1989). The dispositionalist cp-law thus must admit potentially infinite exceptions *internal* to the subject, and the analogy to genuine cp-laws breaks (Kusch 2006, 104-105).

There is a complication that Kusch mentions regarding this last dissimilarity, which is that many psychological laws are indeed similarly cognitively permeable and anyhow sensitive to indefinite, perhaps an infinite number of external defeaters. So, why would cp-conditions not contribute to an illegitimate defence there? Kusch is

not so perspicuous about this point, but he seems to maintain that while genuine cp-laws, at least in psychology, are not meant to apply beyond a reasonable extension of their manifestations, the dispositionalist is bound to claim that Jones' disposition to add remains functional even if he had the brain the size of a universe. In contrast, a respectable psychologist is not bound to claim that her cp-laws would apply to such a subject (Kusch 2006, 105).

In my view, of the four dissimilarities that Kusch discusses, the first is the strongest and decisive, while the fourth is weakest and debatable. In the next section, I shall examine some dispositionalist objections to the first dissimilarity, arguing that they cannot achieve their goal of justifying the use of cp-clauses to support semantic dispositionalism.

3.3 The Return of Semantic Dispositionalism: cp-conditions

Some authors have challenged Kusch's attempt to break the analogy destructive to the sceptic's arguments. Kai-Yuan Cheng (2009) objects to all three main dissimilarities. First, he claims to identify a system of laws to which the disposition to add enormous numbers belongs to:

(SD1) Given an addition problem " $m + n$ " and another " $n + m$ ", both of which involve two large numbers, m and n , I would respond with the same sum in both cases.

(SD2) Given an addition problem " $(m + n) + r$ " and another " $m + (n + r)$ ", both of which involve three large numbers, m , n , and r , I would respond with the same sum in both cases.

(SD3) Given a multiplication problem " $m \cdot n$ " involving two large numbers, m and n , I would respond with the sum " $m + m + m + \dots$ " for n times. (Cheng 2009, 411)

Kusch claims that membership in a merely spurious system of 'laws' is not sufficient for being a genuine cp-law, which is fair. Cheng argues, however, that a system like (SD1-3) is not spurious, only not actual in the sense that no known science employs such laws currently. But this does not imply there could not be such a science; historically, it has taken centuries for the current respectable cp-law systems to develop. The thrust of Cheng's argument against *Systematicity* then is that:

[t]o prove the implausibility of semantic dispositionalism, it would need to be shown whether there are any conceptual difficulties that could prevent any interconnected web of theory from being built in semantic dispositionalism. This Kusch does not do. (Cheng 2009, 412)

A similar argument against *Systematicity* is pursued by Robert Kowalenko:

[T]hese systems of laws or bodies of theory cannot be expected to be known in their entirety at the time of discovery, or to always be discovered *en bloc*. In other words, any requirement that the relevant body of theory be actual—in the sense of being currently known and put forward by actual scientists—is too strong. We can hence dismiss Kusch’s first criterion. (Kowalenko 2009, 186)

I find these responses question-begging. Cheng and Kowalenko claim that the onus is on Kusch to show the *conceptual* impossibility involved in formulating a system of interconnected semantic dispositionalist laws; a demand that effectively means showing the *actual* impossibility of Jones adding with enormous numbers. But Kripke’s original point was precisely that semantic dispositionalism is implausible because we cannot know how Jones would behave if he had the *actual* capacity to add with enormous numbers: the left-hand side enhanced with magic potions et al. cannot justify the right-hand side of producing enormous sums in the conditional. For the same reason, we cannot say that enhanced Jones would *not* produce sums. For the conditional reading of the disposition, the body of laws must be real to determine how enhanced Jones will respond to enormous numbers.

Both Cheng and Kowalenko also argue against the other dissimilarities pointed out by Kusch, and their arguments are more effective here. For example, against dissimilarity from *Approximation*, Cheng accepts that there is a vast gap in how special sciences and semantic dispositionalism may approximate reality with their idealisations, but argues that the gap is one of *quantity* and not *quality*. Namely, while the gap between, say, the Ideal Gas Law and actual gas behaviour is relatively small and convergent, the gap between Jones and enhanced Jones capable of adding with enormous numbers diverges and becomes immeasurable. However, there are also important similarities in the gaps that override the merely quantitative difference: in both cases, the gap is unbridgeable in nomic and conceptual senses. We already noted that it is nomically impossible for the Kinetic Theory of Gasses to become reality because there is no perfect elasticity, molecules always have minimal spatiality, and the effects of electrical forces are not completely neutralised. But this also creates

a conceptual problem because the molecules are constituted by electrons and protons, which arguably have electrical charges *essentially* as their properties, in the sense that there is no world where we can individuate these objects without reference to this property. Even if the charge is not essential to protons and electrons in the metaphysical sense, we do not know what chargeless electrons would mean in the light of our current science. If the antecedent of the counterfactual constituting the Ideal Gas Law was true, the concepts included in the consequent would lose their standard meaning. Cheng notes that the dual unbridgability of the gap between actuality and ideality seems to pose a general problem for cp-laws, and thus does not uniquely disqualify semantic dispositionalism.

Cheng is right to observe that the problem is more or less generic to cp-laws, which can thus be understood in two different ways. A cp-law is *idealised* if its content could not *actually* be true or is very unlikely to be true of all the actual cases. Take the Müller-Lyer law in psychology as an example, which states that humans perceive two lines of equal length differently, depending on whether they end in 'feathers' or 'arrows'. Assuming that it is a genuine law, any data we gather will still most likely be contaminated by external, contingent, and unsystematic interferences, which is why the law requires a cp-clause. However, there need not be any *conceptual* reasons that some dataset could not perfectly align with the predictions of the law, which is not the case with cp-laws as *abstractions*. The Ideal Gas Law is an abstraction in the sense that if the antecedent of the counterfactual was true and there were chargeless, non-spatial protons and electrons, the meaning of the concepts on the consequent side would be changed because molecules that are constituted by the protons and electrons would lose their essential properties – or at the very least, the properties that our current science understands to be essential. The funny thing about the Kinetic Theory of Gases, of which the Ideal Gas Law is a part, is that it works for prediction and explanation, although the truth of the antecedent claim would lead to a conceptual contradiction.

Considering these and other reasons, there is some truth to the analogy that Cheng, following Fodor, pursues. (Indeed, rare is the analogy, which would have no truth to it.) The central common denominator between special sciences and semantic dispositionalism can be located, from my understanding, within the familiar problem of the underdetermination of theory through empirical evidence. Typically, science

works by gathering data, identifying regularities in it, and explaining the regularities by a hypothesis, although not necessarily in this order. (Strictly speaking, what is explained is not the data, but the phenomenon that underlies it, but that point can be ignored here.) The problem of underdetermination is that no finite dataset can logically confirm a general hypothesis, only potentially defeat it, because the data by itself does not exclude all alternative hypotheses. At a glance, the situation looks very similar to the problem facing semantic dispositionalism: the hypothesis is the meaning sentence 'Jones means addition by '+' and the 'data' is his disposition to add with enormous numbers. The dispositionalist admits we require a cp-clause to support the hypothesis, for we know that no one actually has a brain capable of adding enormous numbers; essentially an idealised reading of the cp-clause. However, nothing conceptually seems to foreclose the possibility of such a brain, which is the observation Cheng starts with. And even if there was, that is only to change the reading of the cp-clause from an idealisation to an abstraction, both of which are regularly used in special sciences. So where is the problem in the analogy?

My view of the situation is this: although Fodor, Cheng, and Kowalenko are right to note the generic similarity between semantic dispositionalism and special sciences regarding the underdetermination problem, they miss the more important dissimilarity in the respective *order* of prediction and explanation. Special sciences seek to predict and explain the data (read: the phenomenon) *by* the hypothesis, which always remains defeasible because of contingent external interferences or because of genuine exceptions to the law. In contrast, semantic dispositionalism seeks to explain the truth-conditional *meaning* of the hypothesis ('Jones means addition by '+') by 'predicting' how he *will* or *would* behave with enormous numbers. The sceptic asks what determines that the prediction is accurate, which is where the cp-clauses are employed following the model of special sciences. However – and here comes the rub – the *idealisation* reading of the cp-clause corresponds to the *conditional* reading of the disposition, which, with semantic dispositionalism, proves empirically unjustified, for we do not know *in the light of our current sciences* if there is any determinate way how enhanced Jones would behave with the '+' sign. On the other hand, the *abstraction* reading of the cp-clause corresponds to the *counterfactual* reading of the disposition, which proves to be circular insofar as it merely assumes that Jones' intention is to add. The 'prediction' made by dispositionalism is thus either empirically unjustified or circular; hence, it cannot support the truth of the

hypothesis that Jones adds or the meaning, which the dispositionalist wants to give to the meaning sentence. Whether we read them as idealisations or abstractions, the cp-conditions do not essentially add anything new to the original options of reading the disposition as conditional or as counterfactual. The bottom reason for this is that semantic dispositionalism as it stands has no *actual* (i.e. being *currently* put forth in a system of interconnected laws used in explanation by some special science) empirical laws, which could be legitimately supported by the cp-clause. If we give up the idea that the empirical laws have to be actual, i.e. assert that the onus is on the sceptic to show that there *cannot* be such laws (of Jones adding with enormous numbers), the dispositionalist claim defended becomes trivial and the cp-clauses redundant, for we do not need cp-clauses to argue that it is *possible* (in any sense) that enhanced Jones adds and not quadd. What they are essentially needed for is the claim that he *will* (in all possible circumstances) and *would* (in all counterfactual circumstances) add and not quadd. Unfortunately, that claim they cannot support either, for we do not know what enhanced Jones would actually do because (a) we don't know what the idealised conditions would in reality entail, and (b) we have not been shown a robust fact to determine that his abstracted condition would individuate the intention to add and not quadd.⁸

I think it is in part because they miss the true construal of the original semantic dispositionalist claim that Fodor, Cheng, and Kowalenko are led to their analogical argumentation. They are right where there is an underdetermination problem at stake on both sides of the analogy, to which cp-clauses can be used as a partial solution. However, as I already explained, the key dissimilarity (which Kusch does not quite make explicit enough) is not with the underdetermination relation as with the order of explanation and the 'direction' that underdetermination takes (i.e. the relative position of relata). The truth-conditional meaning of 'Jones means addition by '+' in semantic dispositionalism cannot be justified over the sceptic's alternative hypothesis, leading to its failure. It is because the objections change the claim from actuality to the possibility of adding enormous numbers that the rest of the

⁸ This response to the objection by Fodor, Cheng, and Kowalenko is essentially the same as what Kripke already foresaw: "Let no one – under the influence of too much philosophy of science – suggest that the hypothesis that I meant plus is to be preferred as the simplest hypothesis. [...] Now simplicity considerations can help us decide between competing hypotheses, but they obviously can never tell us what the competing hypotheses are" (WRPL, 38).

similarities that they name between semantic dispositionalism and special sciences regarding cp-clauses are off the mark.

There is one more equivocation that should be settled here before moving on. Is it possible that a confusion takes place between the *a priori* and *a posteriori* construal of semantic dispositionalism?⁹ The *a priori* construal aims to analyse the *concept* of meaning dispositionally, whereas the *a posteriori* construal aims to identify the *property* of Jones meaning addition by '+' in his dispositions. It could then be claimed that Kripke and Kusch successfully show that cp-clauses cannot be used to save the *a priori* construal, which can however support the *a posteriori* examination of the dispositional property of meaning something determinate by a word. Granted that there is no actual science that would show the truth of the *a posteriori* construal, it is surely possible that there will be someday. 'Why, of course,' agrees the sceptic. 'I admit it to be *possible* that science will one day show Jones to be an adder (or a quadder) even with enormous numbers. In the meantime, can you show me what facts determine that there *will* (in the actual world) or *would* (in some ideal conditions) be such a science and a system of cp-laws that shows Jones can behave determinately with enormous numbers? No? Well then...' In short, the debate is not about what hypothetical science *might* show to be true and factual in the future, but about what current science can show to be true and factual right now.

3.4 The Return of Semantic Dispositionalism: abstractions

Other authors have sought to remedy the dispositionalist account by relying either on idealisations or abstractions. Adam Podlaskowski's (2012) defence of semantic dispositionalism operates on two key distinctions: idealisation/abstraction and competence/performance with a word. He sees the problem with idealisation, which Kripke and Kusch point out, but argues that where idealisation fails, the abstractive reading of cp-clauses, together with a focus on our competence-constituting dispositions to use words, will solve the challenge. The idea is that Jones has two kinds of dispositions: those that determine his competence with using a word correctly, and those that determine his actual performance in using it. Podlaskowski

⁹ Scott Soames (1997, 232) tries to refute the sceptic on these grounds.

agrees that the performance-dispositions cannot be saved by cp-clauses, for the familiar dilemma Kripke discusses. However, he claims we can abstract away the performance-dispositions, i.e. ignore the limitations of Jones' finitude in *processing* the enormous numbers and arithmetic operations. What then remains are the competence-constituting dispositions of Jones to use '+', which are no longer exposed to interference resulting from his finite *performative* nature.

But how is it determined *which* competence-constituting disposition is abstracted: addition or quaddition? Podlaskowski writes:

Key to establishing that the agent in the [abstracted] model (and by association, the actual agent) is disposed to add is that the relevant combinatory dispositions are factorizable, just like dispositions to produce a sum within any given column of an addition problem. As such, the relevant combinatory dispositions constitute the basic sorts of coordinating tasks between columns in the order indicative of addition. Accordingly, the agent in the model (for whom the limits of the actual agent are ignored) is disposed to act in the *same* way within any given column and in the *same* way (and order) moving from column-to-column. That is, for an agent without limits, the specified factorizable dispositions yield the set of dispositions indicative of addition. (Podlaskowski 2012, 463-464)

By 'factorizable' dispositions, he refers to Neil Tennant's (1997) proposed solution to the problem of finitude. Tennant claims that Jones' total dispositional state is not required for solving addition problems with enormous numbers because a calculation problem of any length can be divided into manageable sections. The solution to any addition problem can be constructed by iterating the process of solving the finite sections, a capacity that itself needs only to appeal to dispositions. Kripke discusses this solution under the 'algorithmic response' (WRPL, 15-18). Kusch criticises Tennant's proposal by pointing out that because of Jones' finite nature, he cannot execute the algorithm procedure an infinite number of times, which is what it takes to solve enormous equations; essentially, the same problem recurs at a different level (2006, 132).

Does Podlaskowski's proposal fare any better? I shall argue that it does not. For one, he seems to think that underdetermination concerns only the performance side of competence/performance distinction. Assuming we can abstract away the performance dispositions only, something that itself is not self-evident as Podlaskowski acknowledges, the sceptic may yet ask what fact determines that what

is left are competence-constituting dispositions to add and not quadd. It is true that the agent abstracted free from the burdens of finite performance will go on *in the same way* as his competence with the word (i.e. his intention) determines. This is exactly what Kripke allows the counterfactual reading of the disposition to entail. Whether the 'same way' means adding or quadding is not something the abstracting manoeuvre itself can settle, only assume, leading to the circularity objection. While the abstracted agent is indeed capable of hypothetical infinite uniform performance, what this effectively means is that we cannot then appeal to the hypothesised results of the performance to individuate the competence/intention, for the assumed competence/intention is the sole ground for determining the results – cue circularity.

Podlaskowski recognises the problem that abstraction itself cannot settle what is abstracted, but he more or less sidesteps the issue in a footnote (2012, 466). Assuming that it is *a priori* what we mean by our words, i.e. which function our abstract competences/intentions we individuate, the purported solution becomes more workable. The problem is that the *a priori knowledge* assumption is clearly part of the MD_{LB} package (implied strongly by Privacy and Immediate Knowledge) as Kusch reconstructs it and what the sceptic puts into question *in toto*. As I already mentioned, relying on the *a priori* knowledge of meanings and intentions only obliges the sceptic to rephrase the challenge by asking what facts determine that the knowledge is about *addition* and not *quaddition*. Note that it is not the very possibility of *a priori* knowledge, which the sceptic doubts but the deterministic way its object is supposed to be individuated.

3.5 Semantic Dispositionalism without cp-conditions

At this point, I hope to have shown, together with Kusch, the deep problems that haunt any dispositionalist attempt to solve the problem of finitude by employing cp-clauses. Indeed, some self-identified semantic dispositionalists themselves are rightly sceptical about the *ceteris paribus* strategy, although they do not thereby abandon the whole ship. I will discuss two recent papers that question the use of *ceteris paribus* clauses as a solution for dispositionalism in this and the next section. The strategy of both Arvid Båve (2020) and Jared Warren (2020) is that the sceptic paints an unrealistic and ultimately indefensible picture of semantic dispositionalism.

Interestingly, while Båve thinks the picture offered of dispositionalism is much too complicated, Warren says it is rather too simplified. Nonetheless, I shall continue to argue that semantic dispositionalism in either of its renewed forms fails in solving the problem of finitude.

Båve begins by re-articulating the central claim of semantic dispositionalism, which he takes to be implausibly characterised by Kripke as implying that meaning sentences should be truth-conditionally analysed as dispositions to apply the concept up to potentially infinite extensions:

It is as implausible as taking meaning *prime* to consist in a disposition to tell, of each number, whether it is prime. Meaning *prime* is more reasonably identified with a disposition to accept something *obvious*, which can be seen as a definition of “prime”, and similarly for meaning *plus* by “+”. (Båve 2020, 1756)

By 'something obvious', Båve follows Christopher Peacocke's (1992) suggestion that the meaning of 'plus', or the conditions of possessing the concept, means having a disposition to accept its definition. His new solution differs importantly in detail, however, e.g. in eschewing the use of cp-clauses and the normal or ideal conditions for accepting the definition, which ends up folding back to the original dilemma (Kusch 2006, 130). Hence, I will not say much about Peacocke's proposal here, focusing solely on the new version.

Båve takes as his paradigmatic case the possession conditions of conjunction, which he thus describes:

(D) x possesses AND just in case, for some concept c ,

(i) for every proposition p, q , if x were to (a) consider inferring p (or q) from $f(c, p, q)$ for an appropriate amount of time, and (b) have no motivating reason against this inference, then x would immediately infer p (or q) from $f(c, p, q)$

and

(ii) for every proposition p, q , if x were to (a) consider inferring $f(c, p, q)$ from p and q for an appropriate amount of time, and (b) have no motivating reason against this inference, then x would immediately infer $f(c, p, q)$ from p and q . (Båve 2019, 1757)

(i) and (ii) give the elimination and introduction rules for conjunction, respectively. The account belongs to inferential-role semantics, which is recognisably a form of

semantic dispositionalism, although Båve is not committed to inferentialism for all concepts. His initial assumption is that if (D) can be made to work for conjunction, then something similar can be worked out for other concepts too in a broadly dispositionalist and inferentialist spirit. Here, I shall accept these terms and seek to show that (D) does not suffice as a case study of refuting the sceptic.

The major hurdle for (D) to jump over is to deal with the various exceptions that defeat (i) and (ii) thus threaten to stop the disposition from manifesting when it intuitively should, or to make it manifest when it should not. What makes (D) stand apart from the other semantic dispositionalist proposals facing the same problem is that Båve tries to deal with the exceptions retail rather than wholesale, meaning without reliance on cp-clauses idealised or abstracted. As he well recognises, this seems like a daunting task, for as simple as (i) and (ii) are, there will be an indefinite, potentially infinite number of circumstances where the dispositions fail to manifest because of (a) malfunction or (b) the finite nature of the subject, whose dispositions they are. The condition (a) corresponds to the problem of error and (b) to the problem of finitude. Here, my focus remains only on how (D) crosses over the problem of finitude.

The key notion, which drives (D), is 'considering to infer'. Recall that the problem of finitude has two main prongs. On the one hand, there are some propositions or numbers that are too long or complicated for any subject to process as conjuncts before succumbing because of the limits of their finite performative capabilities. On the other hand, even if we somehow ignore the limits to performance, the identity of the competence/intention remains a mystery from a factual point of view because the only (*non-a priori*) way to identify whether the subject adds or quadds (or in Båve's case, operates with conjunction or quonjunction) is by the reference to his performance. The way how 'considering to infer' solves both prongs can be, I think, rendered in the following argument:

Argument from Entertainment

1. In order to make an inference, one has to first consider the inference.
2. In order to consider an inference, one has to *entertain* both its premises and conclusion.

3. If one can entertain both the premises and the conclusion of an inference, one can also consider it.
4. Hence, if one can consider an inference for an appropriate amount of time, one has no motivating reasons against the inference, and there is no defeating neural malfunction present, one *will* make the inference.

There is a rather straightforward way how the Argument from Entertainment (AE), as I call it, avoids both prongs of the finitude problem. For the first prong: any proposition that is too long or complicated for the subject to infer with is also too long or complicated for them to entertain. Båve does not explicitly claim this, but I think it is implicit in what he says. If the subject cannot entertain the proposition, they cannot infer with it to begin with; hence, they cannot *fail* to infer with it. Båve puts the point as follows:

By including the precondition of considering inferences in (D), we can solve the problem of finitude. I will assume, as seems natural, that one can consider an inference only if one can entertain its premises and conclusion. But this immediately entails that, for propositions too complex to entertain, the instances of (i)a and (ii)a do not actually obtain, and so (D) does not entail that possessors of *and* actually make the relevant inferences. Thus, propositions too complex to entertain are not counterexamples to (D). (Båve 2020, 1759)

The account's strategy is radically different from the more mainstream methods of arguing against the sceptic, which I discussed earlier. The main idea of Fodor, Cheng, Kowalenko and Podlaskowski was to use cp-clauses to protect the semantic dispositionalist claim that Jones is an adder even with enormous numbers. Implicit in this form of reply is that the possession of a concept entails, *ceteris paribus*, fluency of performance with applying the concept. But because Båve thinks that possessing a concept (meaning something determinate by a word) should be understood *only* as the disposition to accept its definition, he can claim that:

in order to possess [the concept] PLUS (or mean plus by "+"), one need not be disposed to add very well at all, let alone be disposed to add any two numbers. One need not, for instance, have any idea of how to add double-digit numbers like 68 and 57 (e.g., by adding single digits, "carrying", and so on). (2020, 1759-1760)

Båve can claim this because (D) does not imply that there is any kind of minimum complexity of propositions the subject must be able to entertain in order to count as possessing AND (or a minimum ordinality of numbers to possess PLUS). This is

arguably a strange view about concept possession, and definitely very different from what the sceptic's bread-and-butter cases comprise since it completely decouples the possession of a concept from the capacity to use it. I take it then that Båve would accept that someone who could not consider an addition problem as simple as $1 + 1 = 2$ could still be counted as possessing the concept of addition where they accepted the definition for it.

One may also wonder whether Båve's account makes inferring too watertight, i.e. makes it unintelligible how the counterfactuals in (D) could be false at all, which then would threaten to trivialise them. This is where the specifications of *appropriate amount of time*, *motivating reasons*, and *absence of neural malfunctions* enter the scene. The second important aspect of Båve's strategy, aside from decoupling concept possession and performance, is the claim that *all* the exceptions to (D) can be solved by these three conditions, together with the concept of considering, which saves the trouble of resorting to wholesale cp-clauses. The three conditions also enable him to answer the second prong of the problem of finitude, namely how the intention/concept governing the inferences is to be individuated. Because the set of all propositions is limited by whatever capacities the subject has to consider them, any two propositions which the subject can indeed consider will, together with the three limitations, determine that she will infer *only* as (D) says. Relative to a subset of all propositions is that the subject can consider (call them for convenience 'considerable propositions') she is both sufficiently free from error and from bizarre gerrymandering attempts to make him possess the determinate concept, the cost being to ignore the subject's actual capacity to use it.

3.5.1 Objections to Båve

There are two objections that I shall consider against (D). The first one will be left more on the stage of a sketch because of its reliance on intuitions about what concept possession implies, while the second will be a more straightforward application of the sceptic's reasoning.

One strategy to object to (D) would be to say that it is too far from our intuitions about what possessing a concept means in two ways. First, the whole point of

attributing the concept of, say, addition to someone seems to be partly that it licences inferences to the attributee's performance with the concept. But what is the point of attributing the concept of addition to someone who is boggled by considering $1 + 1 = 2$? I think it would be reasonable to intuit that such a subject simply did not possess the concept of addition. Båve might concede that some kind of an objective minimal threshold for consideration capacities is needed, but practically implementing it on (D) would be another affair entirely, for then concept possession would no longer be solely about individual dispositions. A second intuitive problem with (D) is that subjects with significantly different capacities to consider propositions appear to have distinct concepts, to which Båve could argue that the difference in consideration capacities is comparable to the difference between being a good or a bad adder, albeit stretched to the extremes.

Although it is not fully pursued here, what the first objection strategy shows is that Båve's account does not offer a 'straight' solution to the sceptic since it involves a radical revision of the MD_{lb} picture, namely that possessing a concept entails the mastery of use. However, a successful answer to the sceptic need not be straight, and it would be unfair to assume that anyone could in a single paper convince us that the revision is called for. Hence, I shall grant for the sake of argument that we can completely decouple concept possession from the implications for performance, at least for some concepts, or then that there is some kind of an objective minimal threshold for considering available to supplement (D).

Let us now turn to the more straightforward sceptical objection. I think there is a way to leverage the second prong of the finitude problem against (D) that Båve has not considered, revealed by asking what fact determines that the subject x possesses the concept AND as described by (D) and not the concept QUAND as described by (Q):

(Q) x possesses QAND just in case, for some concept c ,

(i) for every proposition p, q , if x were to (a) consider inferring p (or q) from $f(c, p, q)$ for an appropriate amount of time, and (b) have no motivating reason against this inference, then x would immediately infer p (or q) from $f(c, p, q)$ *except when an enormous number of years has passed since x 's birthday, in which case x will infer only p and never q .*

and

(ii) for every proposition p, q , if x were to (a) consider inferring $f(c, p, q)$ from p and q for an appropriate amount of time, and (b) have no motivating reason against this inference, then x would immediately infer $f(c, p, q)$ from p and q *except when an enormous number of years has passed since x 's birthday, in which case x will infer only q and never p .*

As its cousin quaddition, (Q) is of course nonsense: no one possesses such a concept. But what fact determines this? We cannot appeal to the performance of the subject as evidence of possessing either (D) or (Q), because the empirical differences between them will only manifest when some appropriately enormous number of years has passed since the subject's birth – a condition that will never be fulfilled because of her limited capacities. However, the subject cannot apply both concepts simultaneously because they have mutually incompatible consequences in *some* logically possible cases, just like Jones cannot be both an adder and a quadder despite the difference never manifesting in his behaviour. Where concepts are individuated as functions from inputs to outputs, if the inputs for two functions agree while their outputs disagree, they must be incompatible.

The more general sceptical point I am making with (Q) is that although in the literature, the problem of finitude is most often portrayed as concerning only the enormous ordinality of numbers or the complexity of propositions *themselves*, the problem also covers the *circumstances* in which the application of concepts takes place. Going on 'in the same way' can mean tracking either the properties of the referents themselves or the circumstances in which the referents appear. The circumstances themselves can be categorised into those settings, where something about the subject herself or her environment remains constant. The multiple readings of 'going on in the same way' are often ignored because the canonical example of numbers is not very flexible regarding their circumstances of appearance. Yet, as Kripke observes early in WRPL, the problem also concerns those words denoting objects, which have varying circumstances of appearance:

I think that I have learned the term 'table' in such a way that it will apply to indefinitely many future items. So I can apply the term to a new situation, say when I enter the Eiffel Tower for the first time and see a table at the base. Can I answer a sceptic who supposes that by 'table' in the past I meant tabair, where a 'tabair' is anything that is a table not found at the base of the Eiffel Tower, or a chair found there? (WRPL, 19)

Note that it could be the same table talked about in the example, and the sceptic's question would still arise because the (temporal, spatial, cultural, biographical, etc.)

circumstances in which it appears are different. Now, it could be objected that propositions are like numbers in that there are no relevantly different circumstances in their appearance. But what also matters are the circumstances of *considering* the propositions or numbers, not the referents themselves. Where considering is done in space and time, the sceptic can ask whether some circumstances, perhaps the time since the subject's birth or many other qualities about her or her surroundings, in fact is part of the actual 'going on in the same way' that individuates the concept she possesses. Humans, of course, have a known knack for telling what circumstances are relevant for individuating the concept intended, which is not something the sceptic contests. The sceptic only wants to know the *fact* that determines which of the multiple possible characterisations of 'going on in the *same* way' (possessing a determinate concept) applies to the subject – for simplicity, is it (D) or (Q)?

To explicate my argument, constraining the set of propositions to considerable ones fails in solving the *individuation* prong of the problem of finitude because any two propositions the subject may consider can be considered in two qualitatively different circumstances, in one of which it is revealed that she follows (D), the other one in which she is found to follow (Q) or some other gerrymandered alternative instead. Although the considerable-clause manages to suitably constrain the set of relevant propositions so that going on in the same in *that* sense is guaranteed, it does not similarly constrain the set of *circumstances* relevant for determining what has, does, or will count as going on in the same way. The three conditions specifying the sources of failure for making the considered inference do not help because the sceptical argument hits earlier. Båve attempts to define 'going on in the same way' by 'for any two *considerable* propositions...', to which the sceptic adds a question mark in the form of circumstances where the considering occurs.

There are two ways Båve can try to discriminate between the subject's possession of (D) over (Q) (or the other way around) that I can think of. The first is to claim that the case is settled *a priori*. But as we already saw, the *a priori* knowledge of meanings is something the sceptic challenges. So, Båve owes an explanation of how the individuation prong of the problem of finitude is to be overcome before (D) can claim victory over the sceptic. Accepting the definition of the concept is one way to pursue the aim, but Kusch criticises Peacocke on this point, which Båve does not address.

Second, Båve could deny that considering an inference is sensitive to the circumstances of considering. Of course, such a substantial matter is not something to be merely stipulated one way or the other, so an argument is required – but to whom should the burden of making it fall? The case is similar to Cheng and Kowalenko's claim that Kusch needs to prove the impossibility of a genuine system of semantic dispositionalist laws. Here, too, I argue it would be a mistake to assert that the burden of proof is on me, the sceptic, to positively show that (D) *must* be sensitive to circumstances of consideration – ergo, it is impossible that the subject possesses (D) and not (Q). The converse claim Båve would then take himself to be defending is that it is *possible* for (D) not to be sensitive like that – ergo, it is possibly determinate that the subject possesses (D) and not (Q). However, the sceptic's game is 'rigged' in that in order to *legitimately doubt* that there is a fact of the matter whether the subject possesses (D) and not (Q), it is *not* necessary to show the *impossibility* of her possessing (D) or that considering an inference *must* be sensitive to the circumstances of considering. For the sceptic to express legitimate doubt, it suffices to offer an alternative hypothesis that is (a) empirically equivalent, (b) logically possible, and (c) incompatible with the original hypothesis. The target of doubt is not the possibility of (D) (or any other hypothesis) but its status as an *actual* fact. The sceptic's slogan is: if all we know is what is possible, we do not yet know what is actual; and to doubt what really is actual, it suffices to know what is possible.

3.6 Semantic Dispositionalism without cp-conditions II

Another current author who defends semantic dispositionalism without using cp-conditions is Jared Warren (2020). His aim is the ambitious one of giving a straight dispositionalist answer to all three major sceptical problems: error, normativity, and finitude. Unlike Båve, who thinks Kripke misportrays dispositionalism by complicating it too much, Warren claims that a suitably complicated theory of dispositions is precisely what dispositionalism needs against Kripkenstein's 'monster' (2020, 258). Unlike many others, he does not believe that *ceteris paribus* clauses bring the required subtlety (2020, 261). The closest kin of dispositionalism to which Warren's proposal relates is the algorithmic response, which includes subtracting the number of dispositions necessary and sufficient to make Jones an adder. He begins his answer by appropriately distinguishing between two kinds of dispositions:

(SIMPLE) S has a simple disposition to φ in situation C iff S φ s in C directly, not by way of performing any intermediate actions or activities.

(COMPLEX) S has a complex disposition to φ in situation C iff S φ s in C as a result of performing some intermediate actions or activities that S is disposed to undertake in C . (Warren 2020, 262-263)

'Intermediate action or activities' refers to conscious as opposed to subconscious processes. Whereas a simple disposition manifests to the subject herself as a brute response, a complex one includes a certain level of controlled thought in 'going through the motions' of, e.g. an addition problem. The line between simple and complex can change: a habituated adder may answer '125' to '68 + 57' as a simple reaction. Of course, for every limited subject, there will come a point where neither their simple nor complex dispositions will yield the right answer to an addition problem. The original idea of the algorithmic response (WRPL, 16- 18) was that Jones has learned the algorithm to compute the sum of any addition problem, even if he has no actual capacity to 'go through the motions'. The problems here have already been discussed: the algorithm must have been learned based on a finite number of applications of other terms, any of which the sceptic can challenge by gerrymandering, which will cast doubt over the execution of the algorithm itself. As Kripke notes, a successful answer to the sceptic cannot invoke more 'theory' to be interpreted since that triggers the regress of requiring an interpretation to understand an interpretation. The attempts to refine the algorithmic response by Peacocke and Tennant are refuted by Kusch in his work (2006, 130-132). The main point is that since 'more theory' cannot be the solution, some responses by Jones have to be primitive ('blind', as Wittgenstein says) and thus evade gerrymandering by alternative interpretations. The problem there is that the sceptic does not oppose the notion that we find some responses, e.g. to arithmetical problems, intuitively more compelling than others; (SIMPLE) is one way to formulate this option. This will not help to give a straight answer to the dual problem of finitude, however, because there is no guarantee they follow the addition function all the way to enormous numbers. Ultimately, cp-clauses have to be invoked to support the dispositions, be they simple or complex, which I have already argued to fail.

Warren is well aware of the perils of cp-clauses, as he strives to steer clear away from them. Yet, he thinks that the kernel of the algorithmic answer is solid and promises

a straight answer to the sceptic, if algorithms are only understood correctly. To do that, he proposes another, higher-level distinction in dispositions:

(SINGULAR) S has a singular disposition to φ in situation C iff S has a simple (or complex) disposition to φ in C .

(COMPOSITE) S has a composite disposition to φ in situation C iff φ ing is (or would be) the output of the iterated application of S 's linked simple (or complex) dispositions, in C . (Warren 2020, 264)

The key requirement for composite dispositions is that they be 'linked'. The main idea shares a structure with the algorithmic response: the disposition to add is understood as a relation of one type of disposition to a set of other kinds of dispositions. For algorithms, this meant understanding the disposition to add as a complex disposition constituted by a finite set of simple dispositions to perform arithmetic operations. Warren turns the picture around: the disposition to add is not a complex *singular* disposition, but a *composite* disposition that consists of a potentially infinite set of determinately linked simple or complex singular dispositions to perform arithmetic steps. Being 'linked' means that:

upon being given an addition problem, Ludwig is disposed to start with the first step, and then, with the first step completed, he is disposed to move to and execute the second step. He is not simply disposed to execute step 2 in the process in the same manner in which he is disposed to execute step 1. Instead he is disposed to execute step 2 after step 1 has been completed, in this context. And there is a final step, among the linked dispositions, at which point things are concluded. (Warren 2020, 265)

He gives an example for illumination. Consider a set of colour judgements of the form 'colour n looks the same as colour $n + 1$ '. The ability to make this type of judgement is a singular simple disposition. It is by necessity a finite disposition because the subject cannot judge an infinite number of cases. Where the singular dispositions are linked, the subject not only has a finite (simple or complex) singular disposition to make arithmetical steps, but also has the composite disposition that links the individual steps together such that making 'step 1' results determinately in 'step 2', all the way to the final step. There is no singular *actual* disposition that enables the subject to count with enormous numbers, but only a potentially infinite series of finite, singular dispositions appropriately linked that, together, compose the composite disposition to add. Warren writes:

By talking about Ludwig completing a step and then 'moving to' the next step, I'm relying on the fact that Ludwig's dispositions are structurally linked, so that having completed step k , there is a determinate fact about what the next step in the process is. (2020, 265)

It is time to repeat the jarring old sceptical question of what is the 'determinate fact' that guarantees the arithmetical steps, no matter how enormous the ordinality is, would be performed according to the addition and not the quaddition function. How is the identity of the intention governing Ludwig's composite disposition to be individuated? No singular disposition, be it complex or simple, can determine this, for we have seen time after time how singular dispositions cannot guarantee infinite extensions without resorting to unjustified *cp*-clauses. Warren aims to solve the problem by relying on a set of singular linked dispositions that, together, form a composite disposition. The set of singular dispositions has to be potentially infinite because it is needed to answer enormous addition problems. Actual performance is not at stake here, only the identity of the intention is and whether the composite disposition leaves it indeterminate. Note that the composite disposition itself does not determine anything; instead, it is itself determined by the linked singular dispositions (Warren 2020, 267). But what then determines that the potentially infinite set of linked singular dispositions forms the composite disposition to add and not quadd? The clear risk is that if 'being linked' is itself understood as a singular disposition, Ludwig suddenly finds himself inflated with infinitude.

Warren's purported solution for the individuation challenge links with his answer to the problem of error. Indeed, in a sense, the individuation problem just is the problem of error because we cannot deem an answer by the subject as erroneous unless we know which function his dispositions are supposed to track. For clarity, I will approach the issue as the problem of individuation.

Warren has a so-called 'screening' strategy to meet the problem, which means combining three conditions jointly sufficient to individuate addition and not quaddition. The conditions are:

(NORMALCY) A situation w is *normal* with respect to S just in case neither external nor internal factors are interfering with S 's general cognitive functioning in w .

(M-GENERAL) S has an M -general disposition to φ in a class of situations C , if in the overwhelming majority of specific situations w in C , S φ s.

(STABILITY) An answer a to a question is *stable* just in case as the number of independent repeated trials increases, the ratio of non- a answers to a answers tends to zero. (Warren 2020, 271-272)

The conditions are relatively loose by design, for Warren thinks that part of the dispositionalist failure thus far has been to insist on normal situations implying that error is impossible for the subject. In contrast, in a normal, M-general situation where the subject's response to an addition problem is stable, errors are still possible. Otherwise, Warren's core idea for individuating errors follows the dispositionalist spirit: erroneous applications should be understood as deviations from the stable responses the subject gives in the normal M-general situations. 'But,' will the sceptic ask, 'what determines that the deviations are genuine errors and not evidence that the subject tracks a different function than what we had thought?' Warren replies:

Ludwig is computing the addition function and not the quaddition function because his composite dispositions themselves ultimately consist of various singular dispositions, for which we have already solved the problem of error. We should require that all of these composing singular dispositions are themselves M-general and stable over normal situations. So when Ludwig slips up and forgets to carry the '1' in the 17th step of an addition problem, he is making a mistake in computing the sum, rather than correctly computing the quum, because he has a stable, M-general disposition to carry the '1' in the 17th step of this problem in normal situations. If he were computing the quaddition function, he wouldn't have this disposition. So as long as both complex and composite dispositions are characterized in terms of simple and singular dispositions that avoid error, no further problems of error or indeterminacy will arise at a higher-level. (2020, 273)

3.6.1 Objections to Warren

The problem that I think haunts Warren's proposal concerns the question of what determines when the series of linked, simple dispositions *halts*. Even if it is accepted that the subject is linked, simple dispositions contain the counterfactual potential to produce an infinite series of determinate arithmetical operations, it is equally important for solving the challenge that the process has a determinate endpoint, and not only that every individual step in it is determinate. To give a determinate answer to any addition problem, the subject must halt at a certain phase n of the process. N need not be the correct phase, for it is admissible that Ludwig makes a mistake and counts too far or too short; yet, it is necessary that he has the potential ability to solve enormously large addition problems and the ability to halt the process at a

certain determinate phase. That is exactly what Warren claims is the case: '[T]here is a final step, among the linked dispositions, at which point things are concluded' (Warren 2020, 265).

I contend Warren is not justified to hold that the subject's linked dispositions to perform arithmetic operations in a certain determinate order would have a determinate endpoint at which the process halts. The 'halting problem' facing his proposal is therefore inverted in relation to other dispositionalist accounts, the traditional problem of which has been to show how any finite subject could have the potential to solve enormous addition problems. Warren's Ludwig may well be said to have the composite disposition to produce some kind of answer to enormous addition problems, but it is less clear whether he can halt the process at certain determinate phase n .

It might be wondered how on earth halting could pose a problem for any dispositionalist account when *preventing* the halting was supposed to be the tricky bit. To begin with, we must see that the *cause* of the halting matters. Were the subject to halt the process because of his finite lifespan or dozing off, we cannot say that she has therefore given an answer to the addition problem. The reason is the same as what other dispositionalist accounts used to overcome the contingent limits of the subject and their relevance to the individuation of the arithmetic function she is tracking: the contingent causes of halting are not evidence that she would not have carried on the process, were she suitably enhanced. To solve the halting problem, Warren has to show a cause for endpoint n that depends on the linked arithmetic dispositions themselves.

The problem is that the counterfactual disposition itself cannot contain an answer to where n locates in the process in the actual world, i.e. where Ludwig actually halts his calculation, for the counterfactual disposition is purely potential in nature. What this means is that we cannot individuate the endpoint n from the subject's counterfactual disposition because we cannot then refer to her actual behaviour to individuate the meaning or intention governing her use. Since the identity of the function governing the subject's use depends on the phase where she would end the process, to find out the identity of this n , we must refer to something else than the counterfactual disposition itself to avoid circularity. The claim that n is *a priori*

knowable faces the same problems as other appeals to *a priori* knowledge do: by themselves, they do not exclude the alternative sceptical *a priori* hypotheses.

In other words, Warren requires some actual, non-counterfactual element in Ludwig to claim that his composite counterfactual disposition halts at some determinate phase n . As it was seen before, no actual intention suffices. However, neither will any actual disposition do, because for every unique addition problem, there corresponds a unique endpoint n . In order to have the potential to give a determinate answer to any addition problem, Ludwig would therefore have to have an enormous number of actual dispositions to account for the enormous number of endpoints. But no finite subject can have an arbitrary number of such actual dispositions. As a result, Warren cannot deny the sceptical alternative hypothesis that in some region of enormous addition problems, Ludwig's composite counterfactual disposition does not halt in the actual world but would continue forever, giving no determinate answer, making his behaviour with the '+' sign indeterminate.

Despite its ingenuity, Warren's proposal ultimately succumbs to the same basic problem facing other dispositionalist accounts, which is hooking up a finite subject with an infinite function in a determinate manner. Whereas most dispositionalist theories cannot show how the subject could in principle behave in any determinate way in the region of enormous numbers, Warren's challenge is to find a way for the subject to halt the process at some determinate endpoint with each novel addition problem. The original problem prevails, and Kripkenstein's monster has claimed another victim.

4 REFUTING PLATONISM

The refutation of 'Platonism' in WRPL forms a natural extension of the dialogue on semantic dispositionalism. At once, it should be clarified that Plato himself has little to nothing to do with the topic; his name just is currently synonymous with the kind of realist response available to the Low-brow Meaning Determinist. The major 'Platonist' is in fact Frege, whose theory of sense and reference Kripke discusses as a potential response to the sceptic, albeit briefly (WRPL, 54).¹⁰ For Frege, as for certain types of mathematical realists, the addition function is an objective abstract entity that exists independently of actual, finite minds. Finite minds have mental 'ideas' that are associated with linguistic signs like '+'. The minds are said to *grasp* the senses (equally abstract and objective as the mathematical functions) responsible for determining the addition function as the referent of the plus sign.

The basic problem with Frege's theory, which Kripke mentions, is that it is unclear how a finite subject's mental state (the idea) can grasp the sense capable of determining a unique extension for the plus sign. What fact determines that the subject grasps the right sense and not, e.g. quaddition? This is essentially the same problem as with dispositionalism, but Platonism, so defined, faces another, more familiar difficulty related to grasping. The mental idea is assumedly (Cartesian quibbles aside) a temporal, spatial entity inhabitant of the causal realm, while the sense is an abstract, eternal object that exists outside space, time, and causation. The 'grasping' of sense cannot be a causal event because nothing causal can affect a non-causal abstract entity, yet it is unclear how anything non-causal could in reverse affect an individual's mental state. Is grasping itself causal or non-causal then? Neither answer seems to make sense. Because Platonism in the Fregean guise faces all the same problems as the dispositionalist response does and some more, only three paragraphs are dedicated to its treatment in WRPL.

¹⁰ In fact, he says: 'If it is so brief that you find it obscure, ignore it'.

Not all have yielded in the face of a seemingly insurmountable challenge. Kusch mentions Jerrold Katz (1981; 1990) as the most ardent defender of linguistic Platonism. Nonetheless, not even his admirable efforts and ingenious revisions repel the sceptic, as Kusch argues (2006, 137-145). The key problem, which I will not repeat in depth here, is how to connect a finite subject with an infinite function in a determinate, uniquely individuating manner. Although in Katz's theory, sense is made finite and altogether more graspable; the addition function remains infinite, so the underdetermination problem is ultimately merely shifted around. Where some disposition of the subject is a necessary part of the truth conditions of meaning sentences, and where Platonism attempts to formulate the truth conditions of meaning sentences, its success is predicated on that of semantic dispositionalism.

It is significant to specify that the problem with 'Platonism' is not simply with the *abstractness* of the realm to which meanings are assigned. David Lewis (1983), whom Kusch does not for some reason include in his critical discussion, has made a comparable proposal that includes objective similarity relations in the world to solve the challenge. The proposal builds on David Armstrong's (1978) theory of universals, which posits that (a) universals exist and (b) they are independent of our conceptual schemes and other systems of classification. Lewis denies that universals *exist* as entities, though he agrees that objective similarity relations hold independently of our conceptual schemes. (The subtle difference is not relevant to study here.) According to Lewis, a class of objects is *natural* iff each member of the class *objectively resembles* every other member. Moreover, classes can be more or less natural, depending on how much the objects resemble each other. If true, this proposal solves the problem of finitude because not all ways of Jones going on will count as 'the same way' of going on. Rather, the unique extensional continuation for a finite series of applications is determined by the most natural objective similarity class compatible with the set. Because natural classes are independent of our conceptual schemes and interpretations, the sceptic's gerrymandering will not work quite like it does against dispositionalism. Furthermore, Lewis does not have to hold that the objective, universal similarity relations are problematically abstract: there is no *prima facie* objection to construe them as ordinary natural kinds, hence as robust facts discoverable by empirical science.

In response to Lewis, John Newson Wright (2012) has offered an application of sceptical reasoning that proves as devastating as the original. In fact, Wright points out several problems with Lewis' proposal, e.g. that it is not in agreement with the many epistemic conditions Kripkenstein puts on any straight solution to bear. The main objection, however, is that, granted that there are objective natural similarity classes, some of which are more natural than others, the issue remains of why Jones *will* (in any actual conditions) and *would* (in any counterfactual conditions) go on tracking *the most* natural similarity class as opposed to the second, the third, the fourth, etc., most natural class. Observe that Wright's sceptic does not challenge the notion of objective natural classes, nor that they can be discovered empirically; not even that there is an objective naturalness hierarchy among them. Wright focuses most on the normative problem that even if some class, say adding, is objectively the most natural way to go on, this does not entail that it is the *right* or *correct* way to go on. In this sense, he buys into the correctness narrative expounded on in Section 2.1.

The question remains, however, of why Jones' dispositions will and would be *necessarily* (in the nomic sense) tracking the most objectively natural class instead of other alternatives. If the matter is determined, at some point, the tale about it must include facts about the subject herself, which then lapses into the problems we have already encountered in discussing semantic dispositionalism. This is why Lewis' proposal can be considered an instance of 'Platonism' in this context: because it strives to solve the sceptical problem by relying on objective, subject-independent facts. The ultimate fault is also largely the same, namely that even if there are abstract entities or natural hierarchy classes of objective similarity, this is only half of the solution because the sceptical problem emerges as soon as we try to hook up the subject with the intuitively right objective extension by deploying robust facts as mediators. By necessity, these mediating facts must be finite because they must be facts *about* the finite subject, for it is her use of language that interests us. Lastly, note that the problem for Platonism is not merely epistemic. The sceptic wants to know what facts determine which objective similarity or infinite function is being tracked by the subject's dispositions, not whether we can ever know the truth of the matter.

I take it that enough has been said about Platonism to dismiss the very idea. Since subjects' dispositions have to be implicated in the Low-brow Meaning Determinist

story at some point anyway, the detour to semantic objectivity instantiated in the world independently of subjects is not helpful in theory nor strategically.

5 REFUTING QUALIA

After having extensively criticised dispositionalist and Platonist theories in the previous sections, it is at this point appropriate to remind ourselves of the purpose of doing so. The sceptic's main challenge remains to give justifications for the ascriptions of meaning sentences considered as descriptive propositions with truth conditions. Semantic dispositionalism and Platonism represent the main contenders in the *externalist* branch of justification theory because their strategy is to let truth do most or all of the justificatory work. In order to do that, they need to deliver accounts of the truth conditions of meaning sentences. However, we have just seen how difficult the task has proved to be. Therefore, the tentative conclusion is that externalism cannot justify the ascription of meaning sentences because there are no determinate truth conditions for meaning sentences.

It follows from these considerations that traditional externalist methods of justification, such as reliabilism, cannot solve the challenge. For although no reliabilist holds that meaning sentences need to be true in every case in order to be justified (rather, it suffices that we have enough evidence that Jones is a reliable adder and not a quadder), the proposal fails because the very intelligibility of dispositionalist truth conditions is shown to be suspicious. If we cannot speak of meaning sentences as being true or false (in a substantial, robust sense of truth) at all, we can neither talk about the reliability of Jones as an adder simply because reliability means that at least some of his responses to addition problems are *true*.

The major problems of the externalist method of justification dependent on the truth conditions of meaning sentences lead us to review possible internalist alternatives. Perhaps it is a mistake to try to describe the truth conditions of meaning sentences solely in the language of dispositions and objective similarity classes.

What else than externalist truth might the internalist view appeal to as the justification of meaning sentences and the acts of ascribing them? One option

Kripke briefly discusses is the qualia response (WRPL, 41-42). In a nutshell, the suggestion is that if Jones means *addition*, he is *guided* in his use of '+' by a phenomenal state with a specific qualitative profile. Kripkenstein soon rejects this proposal, however, for three main reasons. For one, it does not appear that using words meaningfully is ordinarily accompanied by special, individuating qualia. Second, it is unclear how subjects can recognise a token phenomenal state as of a type with another, earlier token while possibly being in error about the identification. The problem with qualia is that there seems to be no principle of individuating them except by what seems right to the subject experiencing them, which notoriously transgresses the is-seems distinction necessary for a word to have objective (or public) meaning, i.e. its application to be potentially erroneous. If meaning is public, and if internalist justification via qualia provides the mechanism that individuates meanings, then the individuation of qualia too must meet public criteria in order to serve this role.¹¹ Third and relatedly, even if these two problems are waived, it is unclear how any qualia could provide a *reason* to apply a word in one way rather than another in the sense that the speaker *ought* to use it in a certain way and not another. But if qualia are no ground for reasons, how can they be that for justification? Kusch affirms these points of criticism (2006, 21).

Another closely related proposal that Kripke discusses draws from classical, 'old-school' empiricism that can be called the 'picture response'. Here, the meanings of at least some words are determined by associations with mental images, e.g. a 'cube' with that of a cube (WRPL, 42). Virtually the same problems haunt the picture response as do the qualia response, and Kripke notes that the modern reader is unlikely to be moved by this avenue (WRPL, 43). The reasons are to be found not only in Wittgenstein's own discussion of the classical empiricists' doctrine in the *Investigations*, but also in the famous criticism levied by Quine (1951) and Sellars (1956) against the sense-datum empiricist accounts of meaning.

I have not examined the internalist proposal in detail here, mainly because I think Kripke has it right that this method of justification simply is no longer viable to the modern reader. Where meaning sentences can be justified as judgements, the best

¹¹ This dependency does not entail the reverse claim that having any qualia with a distinct phenomenal profile would be conditional on the qualia being subject to public criteria, a topic which Kripke discusses in (WRPL, 100-104, fn.).

strategy seems to be by their truth conditions as described by externalist means. Failing that, what are we left with? The last strategy of justification that Kripke discusses is called primitivism, to which I shall now turn.

6 REFUTING PRIMITIVISM

The primitivist hypothesis is among the more confounding segments of WRPL, marked already by Kripke's observation that '[s]uch a move may in a sense be irrefutable, and if it is taken in an appropriate way Wittgenstein may even accept it' (51). Despite this tell, four influential commentators on the sceptical challenge – Boghossian, John McDowell, Philip Pettit, and Crispin Wright – argue that primitivism offers a non-skeptical, straight answer to the challenge (Kusch 2006, 207). In radical contrast to the mainstream, Kusch argues that the sceptical alternative solution is a *species* of primitivism, which is thus objectionable to the sceptic only if read as supporting MD_{LB}. To put it shortly, primitivism is, following Kripke, 'desperate' when it admits that (most of) the seven claims of MD_{LB} are irreducible to any robust facts but insists that since our intuitions support them, the claims stand for facts anyhow, hence at least some of them must be primitive about meaning. But it is well to wonder just what kind of fact(s) could truly pass the mustard. In contrast, '[m]eaning-sceptical primitivism is the recognition that since there cannot be meaning-determinist truth-conditions for meaning attributions in general, there also cannot be *non-intentional* meaning-determinist truth-conditions for meaning attributions in particular' (Kusch 2006, 234). What this means is that, in Kusch's interpretation, the idea that there would be semantic truths that could be justified in the metaphysical or philosophical sense is given up entirely in the course of the sceptical challenge and replaced by the alternative sceptical solution. I shall examine the shape of the sceptical solution, including some of its problems, in Section 7.

Likely, no one thinks that all seven major MD_{LB} claims are primitively true about meanings. However, the recent literature has seen the rise of more targeted brand of primitivisms, which pick only certain aspects of the MD_{LB} picture as worth defending. Especially the normativity problem has attracted primitivist solutions of late from authors who accept the correctness narrative. Below, I argue against one such account.

6.1 Refuting Ginsborg's Primitivism

Hannah Ginsborg's treatment of the sceptical challenge has two principal components. First, as she favours semantic dispositionalism, Ginsborg thus objects to the conclusion of the problem of finitude discussed in Section 3, i.e. that there would not be any robust facts determining Jones' state of meaning addition by '+'. Second, Ginsborg thinks that although Kripkenstein's criticism of semantic dispositionalism is inadequate in view of the problem of finitude, she concedes that the problem of normativity poses a problem for dispositionalism. Consequently, she proposes a non-dispositionalist, non-reductive solution to the normativity problem. To anticipate, in my view, Ginsborg's argument for dispositionalism fails largely for the same reasons discussed in Section 3, and that her proposal to solve the normativity problem faces some severe difficulties. The problem in Ginsborg's proposal is the challenge of merging the correctness narrative and the role of normativity in word applications.

First, let us start with a brief review of Ginsborg's avowal of semantic dispositionalism. The basic assumption is very familiar:

'If a series of sounds and marks is implicated in a system of nomological correlations, then, according to the intuition driving [the dispositionalist] kind of view, it has all the context it needs to endow it with meaning' (2011, 157).

Some kind of law-like dispositional state, e.g. an ability to reliably classify objects under a term, is a necessary condition for meaning something by a word for Ginsborg. She does not see the problem of finitude as providing an insurmountable problem for dispositionalism, but refers to answers by Blackburn (1984) and Wright (1984) for a solution. She does not address the complications of these arguments discussed by Kusch (2006), and the same problems concern her view as well. To recall, dispositionalism faces severe problems in trying to explicate the nomological regularities that allegedly constitute Jones' disposition to add. The analogy to dispositions of natural kinds, such as salt or ordinary objects like drinking glasses, fails essentially because for these kinds and objects, we have actual working nomological explanations provided by natural sciences, but it is unclear how to extend the same model to explain the disposition to add, for instance. The idealised conditions required for Jones to be able to add with enormous numbers simply

cannot be substantiated under a current understanding of natural laws, to which a substantial notion of nomic laws surely must refer. The other option of reading the required cp-clauses as abstractions runs into the problem of circularity, for we cannot then appeal to the hypothetical outputs of Jones' abstracted dispositional state to individuate what he means by '+'. Ginsborg merely waives this latter problem, claiming 'that there is nothing objectionable about a 'circular' specification of the disposition' (2011, 160), again on the unwarranted analogy to dispositions manifested by ordinary objects. Neither has Ginsborg in her later writings (e.g. 2018) taken the problem of finitude as a serious problem for dispositionalism, which is why I shall not discuss the matter further.

Instead, what is more interesting in Ginsborg's take on the challenge is her proposal for solving the normativity problem, which she takes to be a genuinely independent issue for dispositionalism, unlike, e.g. Fodor does. Her thought is that there is a primitive normative attitude that provides a necessary condition for meaning something by a word, described by her as an 'awareness of appropriateness' such that the way the speaker is disposed to use the word is in the circumstances the *right* way to use it (2011; 2012). Although dispositions are necessary and sufficient to individuate what was meant by a word, awareness of the application's appropriateness is required to distinguish the application from a brute, automatic reaction. For example, a parrot may have a reliable differential disposition to respond with the sound 'red' to red objects, but intuitively, if the animal lacks *understanding* of the sound's appropriateness, the sound has no meaning in the same sense that the human word has. Ginsborg's suggestion is that the primitive normative attitude provides the needed 'understanding'. It is primitive in not being explicable in intentional, semantic, or rule-following terminology, and further in that the awareness does not 'guide' or provide a reason for the subject to apply the word as she is disposed to. In sum:

I grasp the concept *dog* in virtue of being disposed to respond to (say) Lassie by sorting her with the dogs rather than with the cows, and thus by reproducing representations of barking and tail-wagging rather than mooing, but with the important proviso that, in reproducing those representations, I am disposed to recognize their appropriateness to my present perception. (Ginsborg 2018, 1010)

It should be emphasised that Ginsborg's understanding of the role normativity is supposed to play in relation to meaning is very different from what the justification

narrative posits. Since she thinks that dispositions to use a word are sufficient and necessary to individuate its meaning, the further question of how the ascriptions of meaning can be *justified* does not arise for her, since she would assumedly be open to a purely externalist form of justification. Nonetheless, normativity is supposed to play an essential, 'in virtue of' the role for meaningful states, not because it would guide or otherwise justify particular ascriptions of meaning, but because normativity provides the 'understanding', which she thinks is necessary for genuinely meaning something by a word. She also thinks there is support for this reading of the normativity of meaning in Kant's theory of aesthetic judgement (2018).

On the surface, however, it is difficult to see why exactly the further notion of 'understanding' is needed for meaningfulness if dispositions are deemed necessary and sufficient to individuate meanings. For one, we may grant the intuition that an animal like a parrot does not 'understand' the sound 'red' as ordinary English-speaking humans do (Ginsborg 2011, 237). However, establishing that the intuitive difference in understanding implies a difference in phenomenal awareness and experience is another claim entirely, and not well argued for by Ginsborg. Who is to say the parrot does not have a comparable 'awareness of appropriateness' once it has been trained to gawk 'red' on sight of red things? The epistemic justification for either affirming or denying that claim are nowhere discussed by Ginsborg, and the principled difficulties in the attempt should be obvious. The problem here is not simply to affirm that the parrot (likely) does not have a *similar* phenomenal awareness as an average human does (though how could we actually compare them?), being different species, but in showing that whatever phenomenal state the parrot most likely is in gawking 'red' is *of the wrong kind* to constitute *understanding*. In order for that to be true, it would have to be established that there is only one kind of phenomenal awareness necessary and sufficient to constitute understanding in Ginsborg's sense, for to say that the parrot has *no* phenomenal awareness while gawking 'red' seems out of the question. But then the question arises of what justifies the human phenomenal awareness of appropriateness as constitutive of meaningfulness that the parrot's phenomenal awareness lacks. In other words, although the state of awareness itself need not serve a guiding or justifying function, if the state is necessary for meaningfulness, something about the quality of that state must justify it as special among other possible phenomenal states, yet Ginsborg does not explain what that might be.

One way for Ginsborg to defend her account is to emphasise the primitive status of the special awareness. Saying that awareness of an application's appropriateness is primitive means Ginsborg does not have to explain why a specific phenomenal profile is necessary for meaningfulness. It is simply a brute fact of the world that humans ordinarily have the disposition to experience the right awareness, while parrots do not. Her account would then be a version of MD_{LB} primitivism, though with normativity severed from justification.

Like most answers that appeal to primitivism, this too is in a sense irrefutable. It marks the end of justifications by denying there would be reasons for or against the truth of the claim at stake, namely that while a normal human's utterance 'That is red' can be described as meaningful, a parrot's endowed with an identical disposition to use 'That is red' cannot be.¹²

One reason to be unsatisfied with the primitivist proposal is the (somewhat tenuous) principle that only justifiable claims can justify other claims. Certainly, an *un*justified claim cannot justify anything. However, it is less clear whether a primitive claim that is neither justified nor unjustified can justify further claims. Why this point matters is that we tend to take the descriptions of the parrot's and of the human's usage to justify different claims, e.g. that the human, but not the parrot, said of something that it is coloured. Whether the parrot genuinely *means* something by 'red' or not is intuitively taken as the *reason* for denying or affirming indefinite further claims about its gawking. However, if the description of a voicing's semantic status itself is not justified but primitive, we may enquire how possibly any of the claims derived from this description can themselves be justified.

Of course, the primitivist could argue that none of the claims entailed by the description of the parrot or the human are indeed either justified or unjustified, i.e. the entailments of the descriptions of meaningfulness are primitive as well. But then there will be no *practical* difference between the differing descriptions of the parrot and the human except what concerns their semantic dispositions. If the difference

¹² Of course, most actual humans have far richer dispositions regarding the use of 'red' than any actual parrot does since humans can also make explicit inferences with the word. Here, it is assumed for the sake of argument that the dispositions to use 'red' could be identical without the status of meaningfulness being since the parrot lacks the primitive normative attitude.

between describing the voicing of 'red' as meaningful or not meaningful does not itself serve as a reason for further claims, it becomes unclear what practical difference it makes to describe the voicings in semantic as opposed to non-semantic terms, and not merely in dispositional ones. The problem is that intuitively, we make a practical difference in describing a disposition to emit a sound in semantic or non-semantic terms, and that relinquishing this difference or giving a radically new interpretation of it seems very *ad hoc*.

A final, although I think less critical, issue for Ginsborg can be gleaned from Kripke's observation in WRPL, mentioned in Section 5, that ordinarily, it does not seem that our use of words is accompanied by qualia with a specific phenomenal profile. Indeed, it appears possible for someone to use a word they've heard used meaningfully but don't know its meaning, thus lacking, at least in an intuitive sense, 'an awareness of appropriateness' regarding how one uses the word. For example, assume that Jones overhears a conversation where the word 'diaphanous' is used. He conjectures that it means 'transparent', though he is very unsure about it, and later decides to test the conjecture by uttering "That glass is diaphanous" before an audience he takes to be authoritative, seeing how they react. Does he need to have a special 'awareness of appropriateness' of the word's meaning in order to use it meaningfully?

Generally, I do not take intuitions such as this to be solid ground for arguments one way or the other, and there are a few ways to circumvent the issue entirely (e.g. the qualification, which Ginsborg actually makes, that meaningful application is conditional on *potential* awareness of appropriateness), but in combination with the other points raised above, I believe there is sufficient reason to turn down Ginsborg's proposal for primitive normativism in its current form. The main reason is that it is unclear what problem exactly the proposal solves since, due to the appeal to primitivism, there appears to be no way to justify why a certain type of phenomenal awareness is constitutive of meaningful application and another is not.

In a more recent paper, Ginsborg (2021) has diverted from the correctness narrative, which she apparently accepted in her previous work. Whereas the correctness narrative takes the normativity problem in an atemporal sense, Ginsborg draws attention to the pervasive temporal language in which Kripke puts the original

problem. In particular, the main issue is not the atemporal one of why Jones ought to answer '125' given that he means *addition* by '+', but the temporal one of why Jones ought to answer '125' given the answers he has been giving until now. Why is it that '125' and not '5' is the right way to go in the same way as he has been going in the past with '+'? In other words, what justifies the tensed semantic prescription that '125' is the correct answer, or the answer Jones ought to give, given his past usage, meaning and intentions regarding '+'? Although Ginsborg, in her paper, which aims primarily to be an exegetical reading of WRPL, takes justification only in the internalist sense, it is clear that her new reading is much closer to the one I have also supported here.

That being said, Ginsborg's new proposal retains much of her previous position, most importantly, (a) the proposal of primitive normative attitudes as necessary for meaning something by a word and (b) the dispositionalist account of the truth conditions for meaning sentences. I think that at least the first commitment, which he also ascribes to Wittgenstein's *Investigations*, is clear in the following quote:

Not only does Wittgenstein not see a problem with the fact of our having meant something determinate by '+2', he thinks there is a perfectly innocuous sense in which we can mean each and every step in advance. He makes this clear in a further response to the same interlocutor at [PI] §187: 'When you said 'I already knew at the time [that he should write 1002 after 1000]' that meant: 'If I had been asked then what number he should write after 1000, I would have answered 1002'. And that I do not doubt'. This suggests an account of meaning or grasp of a rule in terms of the disposition to regard certain particular ways of going on (say, writing 1002 rather than 1004 after 1000) as appropriate. (Ginsborg 2021, 14)

It is unclear whether Ginsborg here ascribes to Wittgenstein the dispositionalist view that there is a determinate fact, not only about what the subject would *take* herself as appropriately doing with writing '1004' after '1002', but also a determinate fact of what she *would* do in such a situation. On the one hand, it seems questionable whether the fact (if there are such facts) that the subject takes herself as appropriately writing '1004' is sufficient to constitute a determinate fact about what she in fact means, not simply what she takes herself to mean. After all, might she not make an error and take herself to mean something else than what she actually does? On the other hand, ascribing a dispositionalist theory about the truth conditions of meaning sentences to Wittgenstein goes radically against the reading of WRPL endorsed here, and Ginsborg does not, in her paper (or elsewhere as far as I know), provide evidence

of how WRPL could be so mistaken in this respect. That being said, since this work does not systematically engage with exegetical questions concerning Wittgenstein, I let the matter be as it may.

Regardless of what the correct interpretation of Wittgenstein is, I have already given some reasons to think why primitive normative attitudes do not explain why words become meaningful. Ginsborg argues Kripke overcomplicates the role normativity plays for Wittgenstein in relation to meaning by saying that Jones' response '125' must be justified in relation to his past usage, intentions, and meaning of '+'. Apparently, she continues to think that whatever meaning actually is conferred on '+' by Jones' behaviour can be explained by recourse to Jones' semantic dispositions. Were that true, Ginsborg would be right to argue that there is no need to ground prescriptive normativity stating that Jones ought to answer '125' in his past meanings, use, or intentions with '+'; it would suffice to say that this ought is indeed primitive in Ginsborg's sense. The problem is that Ginsborg (2021), though she criticises Boghossian (1989) for missing the temporal dimension of normativity, nonetheless follows him in undervaluing the underdetermination of Jones' semantic dispositions themselves, or the problem of finitude.

7 LESSONS OF THE SCEPTICAL CHALLENGE

When we get right down to it, what makes the sceptical challenge a problem? Following Kusch, I have argued that the most general concern of the sceptical challenge has to do with the semantics of meaning sentences, in a somewhat broad sense of 'semantics'. What is it that makes sentences of the rough form 'S means x by 'y' *meaningful*? How are these sentences meaningful compared to all other possible sentences, e.g. ' $58 + 67 = 125$ ', 'Ponies are the best', 'I would like to have a pony', and so on?

I have also argued for a reading, according to which the main argumentative structure of the challenge concerns the *justification* of meaning sentences, in several senses of 'justification'. The primary concern with justification is the sceptic's challenge to justify the ascriptions of meaning sentences as descriptive propositions with truth conditions. Justification is here meant in the philosophical or metaphysical sense as opposed to the ordinary way, which would settle to asking Jones himself what he meant.

Imagine, now, that the sceptical arguments hold and that all the three main non-primitivist truth-conditional strategies turn out to be wrong: meaning sentences do not express propositions with determinate truth conditions. It follows that in justifying the ascription of meaning sentences, what we are, in fact, trying to justify is not their truth understood in the 'robust' sense of actualised truth conditions. So, we are left with either of the two forms of primitivism: MD_{LB}, and the sceptical solution.

Other options have also been proposed. Brandom (1994) and Gibbard (2012), for example, take it that the root problem with the truth-conditional strategy is to understand meaning sentences as *descriptions* to begin with. Instead, they counsel us to see meaning ascriptions as performing a different pragmatic role than description, namely an expressive one, although otherwise, there are important differences in

their expressivisms. Huw Price (2004) suggests differentiating subject naturalism from object naturalism, which focuses on the practical aspects of meaning sentences, as an important concept in this category.

There is certainly something to be said for these proposals, and I shall attempt to put in a word on Brandom in the next two chapters. However, I think it is premature to give up the notion that meaning sentences could play a theoretically fruitful, non-deflated descriptive role post-challenge. It certainly seems that we frequently use sentences like '+ means addition' as descriptive statements. That being said, I also believe that the sceptical challenge gives us definite reasons not to think that the way *how* these sentences describe is best thought of in terms of propositions with truth conditions. Perhaps it is propositions and truth conditions that ought to go overboard our Neurath's boat rather than semantic descriptions.

This section will proceed as follows. I shall begin by examining Kusch's interpretation of the sceptical solution – most importantly, its primitivist tendencies – which I think we have good reasons to be dissatisfied with. The sceptical solution is characterised by rejecting virtually all attempts to theoretically grapple with meanings, facts that ground meaning facts, and the ascriptions of meaning. This is naturally an extreme view, and it is questionable whether Wittgenstein himself (or anyone else) ever held it. Thus, my discussion in Section 7.1 is perhaps best understood as heuristic and preliminary rather than conclusive of anything.

Second, in Section 7.2, I shall provide a brief sketch of what I think is the underlying core structure of Low-brow Meaning Determinism that Kusch's discussion, while thorough, does not quite bring to light. What I take to be perhaps the most important lesson of the sceptical challenge is that the meanings of our expressions cannot be *absolutely determinate*, meaning they cannot be determined in the 'infinite' way that drives the problem of finitude. The challenge is thus to be read as a *reductio* about an absolute determinacy of meanings, facts grounding meaning facts, and the ascriptions of meaning sentences. The rejection of absolute determinacy naturally leads us to endorse a temporal as opposed to an atemporal view of meaning determination.

Third, since my tone has been predominantly critical thus far, I wish to end the chapter with a constructive note. What kind of metasemantic theory should survive the sceptical challenge and the problem of finitude in particular? In Section 7.3, I shall propose that the causal-historical account of reference, as inaugurated by Kripke and developed further by Michael Devitt, Mario Gómez-Torrente, and others, is one such theory. I will also criticise Kusch's claim that the causal-historical account is not a good solution to the challenge. While it takes some adjusting, I believe that once we properly understand the lessons of the sceptical challenge and the causal-historical account of reference, we shall have in our hands a powerful theory to explain how at least some words come to have the meanings they indeed have.

7.1 Kusch and the Sceptical Solution

Kusch's interpretation of the sceptical solution has the following key features, on which I shall focus in the following sections:

- 1.) The sceptical solution is minimally factualist about the ascriptions of meaning sentences.
- 2.) Justifications for semantic descriptions and semantic prescriptions come apart.

First, however, I wish to describe the 'spirit' of the sceptical solution, or what the better picture that it proposes to replace MD_{LB} with comprises in an outline. Wittgenstein himself is often credited with a definite quietist attitude towards semantics and perhaps philosophical theorising at large in his later writings. Brandom, for one, breaks this outlook into two distinct views:

- i.) The pragmatic role of the concept of meaning, its home language game, is the explanation of how expressions are used correctly.
- ii.) There is no uniformly or systematically describable way in which meaning functions as an evaluative standard of use. (2019, 18-19)

I think that the notion of 'correct use' that Wittgenstein (according to Brandom) has in mind comes out well in the examination of assertability conditions below, so I

shall focus on the second point in this first section. In what sense is the sceptical solution hostile to systematic semantic theorising?

There are two reasons worth focusing on here. First, the sceptical solution denies 'meanings' (as properties or as a special class of facts) any explanatory role. MDLB suggests that meaning explains use, while the sceptical solution argues that meaning limits use. This idea is most clearly presented by the 'inversion of a conditional' (WRPL, 94-95). MDLB's picture allows us to infer that if Jones means addition by '+', his answer to '58 + 67' will be '125'. This is kind of a law-like prediction and/or explanation of Jones's mathematical activity. The problem of finitude suggests that such conditionals cannot be used to determine unique truth conditions for meaning sentences. Instead, the sceptical solution reads the conditional from right to left: 'If Jones does not answer '125' to '58 + 67', he does not mean addition by '+'.

According to the sceptical solution, because meanings *qua* properties have no role as *explanans*, there is no use any theory can make of them. The second reason why the sceptical solution is hostile to systematic semantic theorising is that the whole phenomenon of meaningful language use, or more broadly, rule-following, is understood in primitivist terms. Virtually the sole role that the ascription of meaning sentences plays in the sceptical solution is as the linguistic community's tests for agreement about how to go on the same way in particular cases. Two people follow the 'same' rule when they agree on enough particular cases, and that they follow a different rule when they disagree on enough particular cases, and that is all there is to rule-following (WRPL, 90-91).

Some have criticised the sceptical solution on the strong grounds that it is, in fact, incoherent as I shall explain below. I concur with Kusch's viewpoint that the sceptical solution is coherent due to its incorporation of 'minimal factualism'.

I think that the main reason to be dissatisfied with the sceptical solution is not that it is incoherent, but simply that it does not say enough about the topic it addresses, and that it does not give us reason enough to believe that more could not be said about meanings, facts grounding meaning facts, and the ascriptions of meaning. This is arguably the attitude underlining Brandom's (2019) treatment of the solution as 'semantic nihilism', but it is also made forcefully by Stephen Turner (1994). Turner

also recognises Kripkenstein as rejecting a certain metaphysical tradition on meaning while embracing a simplified, indeed 'extremely simple', account about rule acquisition: 'one either "gets it" or one doesn't' (1994, 69). The simplicity is then understood in the reductive vein, as reducing our ordinary notion of rules into community agreement, which is clearly insufficient for theoretical purposes:

Familiar difficulties arise immediately on extending the model beyond mathematics. Many of the terms in a language do not have the kind of clear criteria of application assumed by the model – the rules governing them are, as the phrase goes, open-textured. Admitting that usage does not correspond mechanically to ideal models is, indeed, essential to salvaging the explanatory uses of the notion of rules. Unless rules permit a certain openness of application, it becomes impossible to explain how usages shift in response to novel situations and in novel applications. Novel applications would, indeed, be ruled out by the Kripkean model strictly applied. Where the rule was not a matter of agreement, in Kripke's terms, there is no rule at all. (Turner 1994, 73)

In an interesting and possibly unique turn of interpretation, Turner sees the sceptical solution, as read by Kripke, succumbing to the same generic problem that haunts many of the sceptical challenge's targets – the problem of the sameness of meaning across uses and applications. For the MD_{LB} theories, the problem of sameness appeared as the problem of finitude. The sceptical solution rejects the metaphysical strategies to make sense of sameness *qua* absolute determinacy, but keeps the core idea that the sameness of a rule is incompatible with deviations in use across contexts, which then leads to an otiose social theory:

Kripke's Wittgenstein, then, avoids a defective analogical account of the transmission of rules, but only by a kind of sleight of hand. He disposes of the problem of transmission by defining acquisition and agreement as identical, and treating agreement as a primitive concept, a brute fact. But the guarantee of sameness in transmission turns out to rest not on this brute fact – mistakes are possible, if not endemic – but on powers of the community to test for deviance. The tests become the substitute for the common quasi-objects of the other theories – they have the function in the theory of assuring the preservation of sameness. But, like the common quasi-objects of the other theories, such as shared practices, the tests are *themselves* analogical objects, whose 'sameness' is also analogical. But it is a poor analogy. There is nothing about the 'tests' to which a person's acquisition of rules are put in ordinary social interaction or speech that assures that the 'same' rules are being followed. (Turner 1994, 75, fn. omitted.)

I agree with the reasons offered by Brandom and Turner for rejecting the sceptical solution. For all that, it is important to emphasise *these* reasons for rejecting it, for there is something to be learned too from the solution as we shall see below.

7.1.1 Minimal Factualism

Let us start with *minimal factualism*. Several commentators prior to Kusch took the sceptical challenge as a problem for semantic realism in the dual sense that meaning sentences are found wanting not only in terms of truth conditions, but also in terms of truth values. The idea that the sceptic argues meaning sentences to lack truth values was interpreted as entailing that she must advocate 'projectivism' about this section of language, which is an umbrella term for a variety of alternative anti-realist, non-cognitivist stances, e.g. emotivism, quasi-realism, and performativism (Kusch 2006, 148). This reading has then been used to discredit the sceptical solution, for denying that meaning sentences would be 'truth-apt' (have truth values) is argued to be incoherent (Kusch 2006, 151-155).

The question whether the sceptical problem and/or solution advance projectivism or not is important because it sharply delineates the space left for the shape that semantic descriptions can take. Kusch's contribution here, along with George Wilson (1998), is to argue that giving up the truth conditions of meaning sentences does not entail giving up their truth-aptness, making the dichotomy of (truth-conditional) factualism and projectivism false. Indeed, once the dichotomy is given up, we can see that there are multiple ways one can understand meaning sentences as descriptive without having to implicate propositions with truth conditions in virtue of which they are descriptive. Wittgenstein's idea of family resemblance can be understood as one such characterisation: the term 'description' does not denote a uniform logically specifiable category with necessary and sufficient conditions, but a loose collection of jointly or separately sufficient conditions (Kusch 2006, 163).

The way Kusch thinks that the sceptical solution can be understood as declining truth conditions of meaning sentences without endorsing projectivism is by advancing the following three ideas, which I take to form the core of minimal factualism:

- (a) Meaning sentences have assertability conditions,
- (b) A deflationary understanding of truth and facts, and
- (c) A deflationary understanding of propositions.

What are assertability conditions? As Kusch observes, the term is loosely defined in WRPL as 'rough and ready conditions for when it is appropriate, justifiable, permitted or obligatory to make assertions of a certain type' (2006, 27). They correspond only to sufficient conditions, or to necessary conditions qualified with informal vagueness operators, such as 'often', 'frequently', and so on. For example, Kripke describes the assertability conditions for the meaning sentence 'Jones means *addition* by '+' as follows:

Jones is entitled, subject to correction by others, provisionally to say, 'I mean addition by 'plus', whenever he has the feeling of confidence – 'now I can go on!' – that he can give 'correct' responses in new cases; and *he* is entitled, again provisionally and subject to correction by others, to judge a new response to be 'correct' simply because it is the response he is inclined to give. (WRPL, 90)

Furthermore, it is not just meaning sentences but all sentences which the sceptic recommends we look at through their assertability conditions instead of their truth conditions.

Second, to say that 'truth' and 'facts' should be understood deflationally means that we should not think of facts as truthmakers for propositions, or the application of the truth function as predication. Rather, the sceptical solution opts for the redundancy theory of truth, which states that ascribing truth to a sentence just is to affirm the sentence itself, or to deny it in case of ascribing falsehood (Kusch 2006, 172).

Third, a deflationist understanding of propositions means that 'what is said' by an arbitrary sentence should be taken in terms of assertability conditions. We should avoid metaphysically reifying the 'content' that an arbitrary sentence has – the proposition – as something that explains why the sentence is true or false, which in turn requires an explanation of some kind:

'What is said' in an utterance is not determined by some real, natural and universal units of semantic information; 'what is said' is relative to the speaker's and

interpreter's assertability conditions (which will often be shared by many in their linguistic community). (Kusch 2006, 173)

It is difficult to draw together all the implications that these theses imply, but I think that one key lesson has to do with how we should think about the business of 'describing' itself. The MD_{LB} picture thinks of describing, either semantic, empirical, moral, etc., from top down, trying to formulate a three-way logical structure between propositions, truth, and facts that can uniformly explain how the meaningful use of language is possible. It is telling that the standard of 'uniform' or 'systematic' is very strict and has its home in formal logic and theoretical mathematics, which is likely not an accident considering the history of the philosophy of language. The seachange recommended by the sceptical solution instead looks at how we ordinarily apply terms, such as 'truth', 'what is said by a sentence', 'description' and so on, then asks what, if anything, really is shared by all these uses. The result is called 'minimal factualism' (Kusch 2006, 175).

Kusch makes it clear enough that minimal factualism is not meant to be all that can be said about semantic descriptions. As the saying goes, the point is to give a better picture than what the MD_{LB} picture recommends, or a 'framework' to think about what all meaningful sentences share as their minimal common denominator.

7.1.2 Semantic Descriptions and Prescriptions

In Section 2.3, I argued that Kusch's justification narrative differs from the correctness narrative as an interpretation of the argumentative structure of WRPL, while in Section 2.4, I argued that the justification narrative is exegetically and argumentatively stronger. In short, the normativity problem of the challenge is primarily concerned with the justification of meaning sentences in various senses of 'justification'. The most powerful, sweeping argument for why the truth-conditional strategy cannot justify meaning sentences is summarised:

- I. The dispositions of the subject to apply words are a necessary part of any plausible truth-conditional strategy to account for the meaning of meaning sentences ascribed to her.

- II. The dispositions of the subject necessarily can determine only a finite number of actual applications as correct or incorrect, while the truth conditions necessarily determine an infinite number of potential applications as correct or incorrect.
- III. So, it will always remain indeterminate which particular set of truth conditions a given subject's dispositions track, and indeterminate whether she has by any given application made an error or not. But if it is indeterminate when the subject has made an erroneous application, the justification of any application must also be indeterminate.

There are other arguments against the various sub-strategies of the truth-conditional account that attempt to solve this generic problem of finitude, but I think they all fail largely for the same reason. The core problem with both internalist and externalist forms of justification is that the standard of justification – truth conditions – cannot be met by any finite subject, resulting in the incredible conclusion that no ascription of meaning sentence is ever justified.¹³

In this subsection, I shall examine this core deficit of MD_{LB}'s truth-conditional strategy from another perspective, while searching for a way forward. The other perspective has to do with the temporal nature of the challenge.

One important distinction concerning semantic justification, which the sceptical solution emphasises, is the difference between justifying semantic descriptions and semantic prescriptions as mentioned in Section 2.4. One identifying mark of the truth-conditional account is that it sees no difference in the semantic evaluation relative to Jones' past, present, and future applications. The truth conditions of the sentences he is using determine how they should be evaluated in the past, present, and future. The prescriptive issue concerning how Jones ought to apply '+' now and in the future arises only in the metalinguistic sense of whether Jones ought to mean *addition* or something else by '+', which is mostly trivial and allows for an easy instrumentalist solution: Jones ought to mean *addition* by '+' where it will be

¹³ Indeed, in a footnote that has not really been touched upon in the literature, Kripke notes that it is not obvious whether even an infinite subject (i.e. a subject who could have the complete addition table in their mind's eye) could fulfil the standard of justification set by the truth conditional account (WRPL, 52, fn.). I expound on this second way of understanding the problem of finitude in my work (2022).

convenient for others to understand him or to build a house that will not fall over due to the mis-allocated number of nails.

I think it is illuminating to say that the truth-conditional strategy has an *eternalist* understanding of semantic facts. Jones' applications of '+' are correct, if they are correct, in the same sense across all past, present and future uses. The semantic fact obtains in a kind of four-dimensional block-space, which does not discriminate between subjective temporal perspectives denoted by the terms 'past', 'present', and 'future'. The block is homogenous regarding at what moment t the correctness of an application is being evaluated. I think a similar picture can also be found in the background of the correctness narrative inaugurated by Boghossian. As we recall from Section 2.3, for Boghossian, the correctness of an individual application can be raised independently of its relation to Jones' past applications, so that the question of what he means now by '+' is independent of what he has meant previously by it. The question of the sameness of meaning across times can be settled only after we know what it is for individual, atemporally considered applications to be correct or incorrect. Boghossian considers this intelligible because there are perfectly determined facts about how Jones is disposed to use '+' at any given time, which is part of the reason he thinks that the sceptical challenge must be ultimately about something else than the underdetermination of the semantic dispositions themselves. But once we take the problem of finitude seriously and see that there is no way to specify the atemporal semantic dispositions of Jones in a determinate way in and on themselves (i.e. without worrying about their relation to facts about normative correctness), the very idea that Jones' applications of '+' can be evaluated atemporally comes into doubt.

To make that thought clearer, I think we are well served by rethinking the generic underdetermination relation, which formed an important element in the critical discussion of semantic dispositionalism in Section 3. There, I showed that a prominent dispositionalist strategy was to make the underdetermination problem facing semantic dispositionalism analogous with the underdetermination problem facing empirical scientific theories. I argued that the main reason the analogy fails, and why dispositionalism's use of *ceteris paribus* clauses is illegitimate, is that semantic dispositionalism's law-candidates do not count as genuine scientific laws. They do not count as genuine because they are not used in actual scientific explanation of

phenomena, and being potentially capable of being so used does not make them genuine; after all, the sceptical alternative hypotheses are equally potentially capable of being affirmed as actual laws.

Implicit in this understanding of the underdetermination relation is a *spatial* interpretation. The dispositionalist has a finite set of data (Jones' past applications of '+'), which she tries to line up with a determinate meaning-individuating function (e.g. the addition function) considered as a divergent series. Now, any finite set of applications will remain terminally indeterminate in relation to the realm of divergent series, which is to say that any finite set is compatible with an arbitrary, potentially infinite number of such series. Because actuality cannot determine a unique infinitude, the *meaning* of the actual set of applications cannot be individuated by any such divergent series.¹⁴ This is *spatial underdetermination* because we think of meaning itself as already *absolutely* individuated either in the abstract Platonist space or in the concrete, eternalist block-space, only worrying about how to properly 'hook up' any finite stretch of space-time with one 'complete strip' of the block-space. Analogically, if we think of meaning spatially, we could think of a two-dimensional phase-space divided by an arbitrary number of infinite lines or divergent series. Every application of a word, considered as a point, then coincides with some line in the phase-space. But what determines that any point coincides with a unique complete line?

In contrast, if we think of meaning via *temporal underdetermination*, we are thinking of underdetermination of a subject's *perspective* in relation to meaning itself. Here, the difference between the justifications of semantic descriptions and semantic prescriptions also becomes apparent. For the eternalist view of meaning, how Jones ought to apply '+' correctly is determined by how he is disposed to apply it in a potentially infinite number of cases, so that the question between the sameness of past, present, and future applications is at most epistemically interesting. The core idea of semantic dispositionalism is to construe Jones' atemporal state of meaning addition by '+' as *nomologically* related to certain behavioural inputs and outputs. The resulting problems of this assumption were extensively discussed through Section 3,

¹⁴ Not even a disjunctively individuated meaning will do, for assuming that there is no limit to the number of viable divergent series, the disjunction would have to be limitless as well. But a limitless, infinite disjunction cannot *exclude* a word from having one meaning as opposed to another, which is the original problem.

which I think should motivate the thought that the nomological model is not useful in thinking about the relation of meaning to behaviour.

I think the best way to understand the idea behind 'temporal underdetermination' is to realise that meanings have histories, and that their future is not determined by their past. In a sense, the sceptical solution guides us into thinking that meanings *qua* proprieties of use just *are* their histories, as Brandom emphasises:

Many of [Wittgenstein's] thought-experiments concern this sort of process of *pragmatic projection* of one practice into another. We are asked to imagine a community that uses proper names only for people, but then extends the practice to include rivers. There is no guarantee that interlocutors can master the extended practice, building on what they can already do. But if they can, then they will have changed the only ^Sessences^S proper-name usage could be taken to have had. In the old practice it always made sense to ask for the identity of the *mother* and *father* of the named item; in the new practice, that question is often senseless. Again, we are asked to imagine a community that talked about having gold or silver in one's teeth, and extends that practice to talk about having pain in one's teeth. If as a matter of contingent fact the practitioners can learn to use the expression 'in' in the new way, building on but adapting the old, they will have fundamentally changed the 'meaning' of 'in'. (Brandom 2019, 20)

What this *historicity of meaning* means is that there is no simple way to derive predictions about how Jones will (in the actual world) or would (in some counterfactual world) behave with the sign '+' from the description about how Jones has, in fact, used the sign in the past. The facts that ground the meaning of '+' in Jones' use do not form a kind of causal nexus (as Kripke puts it in WRPL, 62) from the past to the future. For the same reason, we cannot be satisfied with the simple solution to the normativity problem, i.e. that the unique thing that Jones ought to do could be derived from the semantic facts of the matter, together with his intentions.

7.2 Temporality of Meaning

In this chapter, I have frequently referred to the 'temporality of meaning' as one important lesson of the sceptical challenge. In this section, I will briefly summarise the ideas implied by the phrase and examine their consequences for the metasemantic theory in the next section.

The founding argument that I have pursued in this chapter has been that there is no 'straight' solution to the problem of finitude, i.e. a solution that would explain how the ascriptions of meaning sentences could express propositions with truth conditions. (Although strictly speaking, a full straight solution would have to satisfy all the main seven thesis of MD_{LB}, since some of these are arguably in contradiction with one another, I take it that a reasonable straight solution will consist of at minimum explaining the truth conditions of the ascriptions of meaning sentences.) The main reason for this conclusion is that none of the four main strategies brought up in WRPL and discussed here deliver the straight solution. My examination has, of course, been limited, but I trust that the case made is strong enough to at least entertain the conclusion that there is no straight solution to the problem of finitude.

What follows is the rejection of what I have called an 'absolute determinacy' of meanings, facts grounding meaning facts, and the ascriptions of meaning sentences. Kusch identifies the same idea under the MD_{LB}'s thesis 'Objectivity' of meaning: "The idea [...] that [the subject's] meaning-constituting mental state contains and determines all future, potentially infinite, correct applications of ['y']" (Kusch 2006, 9). A more definite formulation of absolute determinacy comes from Alexander Miller:

In the case of a descriptive expression such as '+,' whatever fact that is proposed as making it the case that '+' means the addition function must be inconsistent with the hypothesis that '+' means some other function, such as quaddition. In the generalized version of the argument, which applies to both descriptive and non-descriptive language, this becomes: whatever fact that is proposed as making it the case that rule R_a is the rule governing Smith's use of expression E must be inconsistent with the hypothesis that the rule governing his use of E is R_b, where R_a and R_b are such that for some possible use Δ of E, Δ is correct according to R_a but incorrect according to R_b. (Miller 2010, 460)

Another closely related formulation of absolute determinacy comes from what Brandom calls the Fregean determinateness of senses:

Fregean senses are required to determine classes of referents whose boundaries are *sharp*, *fixed*, and *complete*. To say that they are *sharp* is to say that it is impossible for any possible object to fall *partially* in the class determined by the sense (excluded middle), or both to fall in it and to fall outside it (non-contradiction). To say that the referents are *fixed* is to say that the boundaries of the class of referents determined by the sense do not change. (Which sense a given sign expresses may change, if the use of the sign changes, but the senses themselves do not change.) To say that the boundaries of the

class of referents is *complete* is to say that the sense determines a *partition* of the possible candidates: *every* particular is classified by the sense either as falling under the concept it determines, or as not falling under it (excluded middle). This is Fregean determinateness, or determinateness in the Fregean sense. (Brandom ST, 429)

There are differences between these formulations, but I trust the reader can see the essential resemblance, which I think boils down to this: absolutely determinate meanings are atemporally determined. The analogy to eternalist block space continuous to be useful here, I find, since we can think of the subject's semantic dispositions (or whatever fact about her is supposed to 'contain', i.e. individuate or ground, the absolutely determined meaning) as one complete strip of the block, the question being how it is determined that the actual subject aligns with one unique such strip.

From the conclusion that there is no straight solution to the sceptical challenge follows the rejection of absolute determinacy. (To specify, it follows by the simple *reductio* reading that since our actual languages are meaningful, they cannot be meaningful in virtue of meeting absolute determinacy.) It does not follow that we must give up the thinking of meanings as factually determined in any sense, i.e. that we would be forced to accept projectivism about the ascriptions of meaning sentences as explained in the section above. On the other hand, I think we can do better than minimal factualism with the idea of semantic truths, properties, or facts.

What kind of picture about semantic facts emerges in the aftermath of the sceptical challenge? To begin with, after becoming explicitly aware of the idea of absolute determinacy, we can see that the main criterion for semantic facts as understood by the sceptic is that they are something that make the applications (or other kinds of use) of expressions 'correct'. It is important to realise that 'semantic correctness' in this sense captures a much broader idea than truthful application. In effect, the idea of correctness as it appears in the sceptical challenge is supposed to cover all possible semantic values, not just of descriptive language, but also of, e.g. expressive language, as we see in Miller's concise formulation. So, inferential roles, the Kaplanian character of indexicals, verification conditions, etc., can be subsumed under candidate semantic facts that make applications 'correct'. The problem of finitude is

supposed to rebut *all* possible interpretations of 'semantic correctness' in the sense of absolute determinacy.¹⁵

Rejecting absolute determinacy naturally leads us to think that there must be something wrong with semantic correctness itself as understood by the sceptic. Here, a fruitful way to present the problem is made by Jussi Haukioja (2021). According to Haukioja, we can understand the basic metasemantic question (Why does a token expression have the meaning that it in fact has?) in two ways:

(A) What facts about me make it the case that the predicate P (as I use it) denotes a particular abstract entity, A?

(B) What facts about me make it the case that it is correct to apply predicate P (as I use it) in previously unencountered instance I? (Haukioja 2021)

Briefly, question A corresponds to the atemporal way of understanding the metasemantic basic question, whereas B corresponds to the temporal one. To see this, it suffices to recognise that the 'abstract entity' referred to in A just is (at least in my view, though I cannot speak for Haukioja) the *rule* that is supposed to determine a sharp, fixed, and complete semantic value for a token expression. Although the actual semantic values of the expression token need not be 'abstract', the rule that is supposed to *exclude* the alternative gerrymandered semantic values must be. Whatever the semantic value of choice is, it must determine an application (or other usage) as correct in this sense if absolute determinacy is the standard by which meaningfulness of expressions is measured.

What makes question B different is that in order for any actual application to be determined as correct or incorrect, it need not be the case that *all the logically possible* applications of the expression must be already determined as correct or incorrect. The B question is essentially temporal because it asks for determination in a certain particular context where it need not *yet* be determined whether the expression applies to the newly encountered instance correctly or incorrectly. In other words, the B question does not implicate any abstract entity (a 'rule') to mediate the correctness

¹⁵ As we recall from Section 2.1, many have ignored the depth of the problem of finitude, and are thus led to thinking that the truly sweeping argument must proceed by another route, namely that of normativity. In my view, the problem of normativity can only be properly understood against the depth of the finitude problem. I will return to this point in the subsection below.

of the application, for the particular context (including the history leading up to it) is supposed to suffice for settling the matter. (Naturally, it may transpire that some novel instances of application fail to be covered by any existing predicate, so that a new one must be invented.)

It is difficult to overstate how important the distinction between A and B questions is, not just for resolving the sceptical challenge, but also for the broader aims of metasemantics. Moreover, I believe that numerous authors over the years have sought to articulate this difference in various ways, but that it generally has not caught the attention it deserves. To mention just one example, in her review essay of WRPL, Elisabeth Anscombe (1985) observed that Kripke himself apparently had something like the atemporal question mainly in mind when formulating the challenge. Anscombe begins by drawing attention to part of footnote 13 on pages 18-19 of WRPL, where Kripke aims to correct what he thinks is a misreading of the sceptical challenge:

The point (which can be fully understood only after the third section of the present work) can be put this way: If someone who computed '+' as we do for small arguments gave bizarre responses, in the style of 'quus', for larger arguments, and insisted that he was 'going on the same way as before', we would not acknowledge his claim that he was 'going on in the same way' as for the small arguments. What we call the 'right' response determines what we call 'going on in the same way'. (WRPL, 19, fn.)

The key line here is 'What we call the “right” [read: “correct”] response determines what we call “going on in the same way”'. So, the correct application of an expression in a given instance is what determines whether Jones goes on the 'same way' as before. Anscombe regrets this order of understanding the challenge, and in counter to it, quotes the case of the pupil counting the +2 series in the *Investigations*, which ends:

In such a case we might say, perhaps: It comes natural to this human being to understand that order, given our explanations, as we would understand the order 'Add 2 up to 1000, 4 up to 2000, 6 up to 3000, etc.' (Wittgenstein 1958, §185)

According to Anscombe, what can be said here (or what Wittgenstein would say) is not simply that the pupil was not going on the same way as before, but that his way of going on the same way differed from ours (1985, 345). What this seems to imply is that there is a notion of going on the same way that is independent of, or prior to

the notion of a correct application. Perhaps we can take the thought even further: the same way of going on determines the correctness of applications.

In order to develop that thought, we must take on board some further commitments about semantics and metasemantics. The beginning of that task will be made in the final section of this chapter, where I shall propose a few cornerstone ideas that, together, might solve the sceptical challenge, understood in the sense of the B question rather than the (in my view, insoluble) A question. The core idea builds on the causal-historical theory of reference.

7.2.1 On Semantic Justification Again

In Section 2.4, I presented, following Kusch, a rudimentary taxonomy of 'semantic justification' that I took to be mostly implicit in WRPL. The main form which justification takes in the challenge, I claimed, was in the question of what justifies the ascription of meaning sentences in the philosophical or metaphysical sense, i.e. in some more fundamental sense than consulting a dictionary or asking the speaker what they meant. In this section, I want to briefly show what I think is the right understanding of this 'normativity problem' in WRPL in view of the rejection of absolute determinacy.

The central notion that arises from the rejection of absolute determinacy is what David Blood (1997) has usefully called 'meaning finitism':

According to meaning finitism, we create meaning as we move from case to case. We *could* take our concepts or rules anywhere, in any direction, and count anything as a new member of an old class, or of the same kind as some existing finite set of past cases. We are not prevented by 'logic' or by 'meanings' from doing this, if by these words we have in mind something other than the down-to-earth contingencies surrounding each particular act of concept application. (Some interpretive gloss can always be provided to render the step formally consistent.) The real sources of constraint preventing our going anywhere and everywhere, as we move from case to case, are the local circumstances impinging upon us: our instincts, our biological nature, our sense experience, our interactions with other people, our immediate purposes, our training, our anticipation of and response to sanctions, and so on through the gamut of causes, starting with the psychological and ending with the sociological. That is the message of Wittgenstein's meaning finitism. (Bloor 1997, 19)

It may seem that if there are no facts of the matter what would count as the correct application of an expression (say, 'is a dog') in encountering a novel candidate instance of application (an animal named Lassie), there can neither be *normative reasons* (or justifications) for deciding to call Lassie a dog or not. In this simplistic layout, determinate reasons for the application depend on determinate truths of meaning (i.e. what is the absolutely determinate extension of 'is a dog'). Whatever causes there may be to settle whether speakers decide to extend the predicate expression 'is a dog' to Lassie or not are not really worthy to be called their reasons for the application in the sense that the causes would *justify* the application – justify in the philosophical or metaphysical sense, that is.

For all I know about it, this view might be true. Perhaps semantic justifications must be tied to the foundationalist sense in order to be real semantic justifications, and where the foundationalist sense is incompatible with meaning finitism, there is no such thing. At places, it seems that Brandom is motivated to think exactly along these sorts of lines, as we shall see in the next two chapters.

I will not have much to say about semantic justification in this chapter, for the focus on the rest will be on semantic descriptions. That being said, I see no *prima facie* reason to give up the idea that meaning finitism would be incompatible with semantic justifications properly speaking. I agree that our ordinary reasons to call Lassie a dog (she barks, chases cats, has a certain genetic make-up, etc.) are not foundational, metaphysical, and least of all, philosophical. But are they not (normative) *reasons* for all that? Perhaps, following Anscombe, it is possible to understand normative semantic reasons as acquiring their authority not from truth (or other variant of correctness) but from the 'same way of going on'.

Be that as it may, in the last section of the chapter, I shall make some moderate efforts to show how and why we can do better with the idea of semantic descriptions (i.e. descriptions about what an expression token means) than what the sceptical solution with its minimal factualism in Kusch's reading can deliver.

7.3 Causal Cures

In *Naming and Necessity* (1980), Kripke taught us, among other things, how to think about essence as an empirical category. Following Quine's *bon mot* that '[m]eaning is what essence becomes when it is divorced from the object of reference and wedded to the word' (1951, 22), it is a short way to the idea (if not the argument) that some meanings, too, are empirical – at least in the negative sense of not being in the head. In this final section, I want to take that line of thought seriously in the context of the sceptical challenge and solution.

Let us begin by returning to the sceptical challenge in its 'raw' form. Imagine that we encounter an animal named Lassie. Because Lassie barks like a dog, looks like a dog, and chases cats like a dog, we decide to call Lassie a dog. The sceptic who accompanies us accepts the present statement that Lassie is a dog as true, given the ordinary meaning of the word. But then she wonders out loud:

'What makes it true that our word “dog” applies correctly to Lassie, or to dogs in general?'

After some familiar puzzling, we come to conclude that there is no determinate fact as to why 'dog' applies correctly to all and only dogs, as opposed to (say) all the dogs and some cats on Wednesdays. Since the meaning sentence 'Dog' applies correctly to all and only dogs' has indeterminate truth conditions, the sentence 'Lassie is a dog' must also have indeterminate truth conditions because its component word 'dog' does. Hence, the sentence 'Lassie is a dog' will not be either true or false.

To break down this chain to its landmark moments in the order they *appear* to us in the example, we have:

- (1) The question whether Lassie is a dog,
- (2) The question whether 'dog' applies correctly to Lassie, and
- (3) The question of why any expression token in general applies correctly or incorrectly in a determined way.

In the light of the A-question discussed in the previous section, answering (3) is a logical condition for answering (2) and, ultimately, (1). This is part of what it is to understand the sceptical challenge in an atemporal way. In contrast, considering the B-question that rejects absolute determinacy, question (3) ultimately drops out as either unnecessary or bluntly unintelligible. The reason it drops out is that there are no general facts describable by logical rules that would be sharp, fixed, and complete that could answer it. Instead, the facts that determine an answer to (2) must be temporal, local, and 'less than logical'. For all that, though, they need not be completely mysterious.

Let us now take a gander at what the causal-historical account of reference would say about the case. Suppose that in the past, the speakers who encounter Lassie have used the word-type 'dog' to refer to animals much like Lassie. They thus find it natural to count Lassie as a 'dog', simply by so calling him. So, in their consequent use, Lassie is, for all intents and purposes, a dog. But now, something spurious seems to happen. The sentence 'Lassie is a dog' must have been true (if it is at all) long *before* the encounter with the speakers and Lassie; certainly, it could not have *become* a truth when it was determined that Lassie belongs to the extension of 'dog'. But if this was a truth before the encounter with the speakers, then it seems that their decision to extend their 'dog-practice' to Lassie must have been predetermined prior to the encounter in the sense that their decision to apply the word to him can be objectively right or wrong regardless of their views about the matter. (Alternatively, their use of 'dog' might bifurcate in the encounter, becoming ambiguous.) So, it seems that if (1) has a determinate truth value at all, then something like (3) must have an intelligible answer to it after all.

The answer to the puzzle offered by the causal-historical account, at least in a rough form, is that at some point in time, when the general name 'dog' was introduced to the language of the speakers, there was a moment of initial baptism such that the term 'fixed' its reference. Consequently, other speakers not present during the baptism 'borrowed' the word from those who were, starting what Michael Devitt (1996, 164) calls a 'designation chain' (D-chain) for the word-type 'dog'. Further down the chain, the term may experience consequent moments of baptism or 'groundings' that either reinforce the original reference or change it somehow, with or without the awareness of the speakers. So, the case of encountering Lassie can be

read as a case of reinforced grounding of the term 'dog', supposing that the class of referents is the same as with the original baptism, a mistaken application; or the encounter marks a bifurcation of the term's reference, supposing that Lassie is relevantly different from the original stock of referents and that the speaker neglect to perceive this.

The devil of the answer, of course, lurks in the question of what exactly 'same' and 'different' mean in this causal-historical picture of semantic determination. Following Kripke's insight, numerous authors in the causal-historical tradition think that sameness and difference depend wholly on facts about the referents, e.g. what is necessary and sufficient for anything to be a dog. The problem, of course, is that appearances notwithstanding, finding the jointly necessary and sufficient membership conditions in any kind, least of all a biological one, has proved highly difficult and arguably impossible. The risk here is that the reference of many ordinary terms will become indeterminate from a strictly scientific point of view. However, this is different indeterminacy than what Kripkenstein's sceptic trades with, as I shall explain below in connection with discussing Gómez Torrente's take on the causal-historical theory of reference.

Related to the problem of the sameness of kinds themselves is the infamous *qua*-problem, which concerns the determination of our kind terms. It is a slightly bold thing to claim, but I think that part of the reason for what has made the *qua*-problem so nefarious is that various authors in the causal-historical tradition have implicitly accepted something like absolute determinacy as the standard against which the problem must be solved. This is not to say that rejecting absolute determinacy makes the *qua*-problem easier to solve, but it does make it different. For one, understood in the atemporal sense of the B question, the *qua*-problem splits into two parts:

Historical part: What determines which class of referents the D-chains of type 'dog' have fixed in the past?

Futurist part: What determines which class of referents the D-chains of type 'dog' will (likely) be fixed in the future, or ought to be?

My focus in this section will be mostly on the historical part. Below, I shall first examine Kusch's applied sceptical criticism of the causal-historical theory of reference. Second, I attempt to show how the causal-historical theory may be compatible with the sceptical challenge and solution after all.

7.3.1 Kusch's Criticism of the Causal-Historical Response to the Sceptical Challenge

One thing that Kusch thinks to be symptomatic of the causal-historical response's unworkability as a solution to the sceptical challenge is that Kripke never brings it up in his exposition. The main reason it is unworkable has to do with the fact that the causal-historical account of reference fixing is not purely causal, but arguably has to include some intentional elements as well. The basket of problems related here is widely known as the *qua*-problem: *qua* what is, for example, the reference of a proper name fixed in an ostensive setting to the whole object instead of one of its parts or the current temporal slice in which the initial baptism occurs? Or why is it that, when naming a natural kind, such as gold, what is named is the kind gold and not one of the higher taxa that the same sample belongs to, e.g. metal, heavy stuff, etc.?

Kusch is correct that in order to resolve *qua*-problems, referential intentions must be involved in the causal-historical theory. This is indeed something Kripke never contested, as both borrowing and baptising are clearly intentional acts in some sense. However, Kusch thinks that once referential intentions are involved, the sceptical problem wedges in the form of competing sceptical hypotheses how the intentions are to be described. This is his response to McGinn's (1984) effort to refute the sceptic based on the causal-historical theory: in baptising an object with a proper name, the baptiser must specify his referential intention with some descriptive terms ('this *man*, this *mountain*, this *metal*', etc.). But assuming that every term must be so supplemented by referential intentions and descriptions, what determines that by 'mountain', one does not mean *quountain*, which means a mountain or a sinkhole if encountered in a certain latitude or whatever (Kusch 2006, 135).

The only way to avoid sceptical gerrymandering is to appeal to non-descriptive specifications of the intentions in reference-fixing (and borrowing), which is what Penelope Maddy (1984) suggests. Her idea suggests that inspecting the brain states

of the baptiser during the initial baptism can solve some qua-problems. Perhaps the brain states can even discriminate among the higher taxa of the object referred to; assume for the sake of argument that this is possible. Kusch's argument is that even then, we would still require a story of what connects the physical brain state to the mental state corresponding to the right referential intention of the baptiser, or then argue that the mental state is reducible to the brain state somehow. Difficulties with mental reduction aside, Maddy's proposal also faces the problem of finitude, for if the brain states are to actualise the truth conditions of meaning sentences involving natural kinds, it would have to be shown how the brain can determinately classify a potentially infinite number of samples without circularity or an illicit appeal to *ceteris paribus* clauses (Kusch 2006, 136).

Clearly enough, the problems with the causal-historical account, which Kusch identifies, are generic to the truth-conditional strategy as such; largely the same issues haunt the neighbouring Davidsonian idea of intersubjective causal triangulation that he discusses (2006, 83). The question thus becomes: is the causal-historical account necessarily wedded to the truth conditional strategy in the sense that referential intentions (at least in baptism) necessarily need to include some descriptive elements? One should think not, for was it not the original ambition of Kripke precisely to criticise the descriptivist theory of meaning and reference-fixing, which rely on truth and knowledge as key explanans, in formulating the better picture of the causal-historical account? Indeed, I believe that what Kusch's arguments against Maddy and McGinn show is that the causal-historical account cannot deliver a straight response to the sceptical challenge, if by a 'straight' response, one aims to preserve the truth conditional foundations of MD_{LB}. As I have argued above, the challenge is a successful *reductio* against the truth-conditional account of meaning sentences.

What is needed is a revisionary interpretation of the semantics of meaning sentences, starting with how they can be understood as descriptive. I already showed how Kusch defines the minimum threshold for connecting meaningful sentences as descriptions in minimal factualism. I think the causal-historical account is posed to deliver a development from that basis in case of many term types, such as proper names and certain kind terms. The key here is how the referential intentions are to be specified both in the acts of borrowing and baptism.

7.3.2 Devitt on Semantic Descriptions

In this section, I shall discuss a fundamental (or at any rate, important) issue concerning 'semantic descriptions', i.e. descriptions concerning what an expression means, in view of the causal-historical theory of reference. We can begin by posing the following question:

(1.) What is it that meaning sentences describe?

What delineates the question, that we have seen, are two extreme sorts of answers. On the one hand, we have MD_{LB}, which understands meaning sentences as propositions with robust truth conditions; on the other, we have minimal factualism, according to which meaning sentences only have deflated truth conditions. Rejecting the absolute determinacy of meaning sentences also means giving up their truth conditions, for determinate truth conditions must be logically unique in the sense of constituting a sharp, fixed, and complete partition of the logically possible applications of an expression as correct or incorrect.

Michael Devitt (1996) has proposed what I take to be a useful answer to question (1). But what is perhaps even more useful than his answer is the criteria, which he tasks any successful answer to meet. The very first question we should be concerned with in semantics (and metasemantics) is what theoretical role 'meanings' are supposed to serve to begin with. To put it bluntly, why should any theory care to postulate 'meanings' for expressions, or care enough to deny that they can be postulated? It is telling that even though Devitt subscribes to a broadly Fregean approach to semantics, in which the main theoretical tasks of senses is to explain (a) the determination of referents and (b) the cognitive/epistemic 'grasp' subjects have of referents, this alone does not justify the theoretical interest to meanings (by the identification of sense with the relevant nature of meanings). Determining referents and explaining epistemic grasp are, according to Devitt, already too theoretically loaded to meet his basic methodological question.

Devitt's proposals for what should be the methodological basis for postulating meanings are (a) to explain actions of subjects and (b) explain how beliefs are a reliable source of knowledge about the world. But why should we think that meaning

has anything to do with either actions or knowledge? One strong reason that Devitt relies on is the fact that our 'folk practices' of explaining actions and relying on the beliefs of others (or our own, for that matter), for reliable knowledge about the world centre on the ascriptions of meaning in attitudinal 'that-clauses', such as 'S believes that *p*'. In his usage, the predicate 'that *p*' functions as a singular term, picking out an object, which, when the ascription is true, has the meaning-property that explains something about S's actions and serves as a reliable guide to knowledge about the world (Devitt 1996, 56-57).

The object of an ascription of a that-clause is, for Devitt, a token thought; but what is more interesting is the meaning-property that is being ascribed to the thought. For, according to Devitt, at least with proper names, that property is a Fregean sense understood as a causal-historical mode of reference. This is what he has later called 'the shocking idea', which follows from two other, more familiar (and arguably less shocking) thoughts:

Externalism: Some words, including names and natural kind words, refer in virtue of causal relations that are partly external to the head (and hence these words are to be explained by a theory that is not a description theory).

Meanings as modes: A meaning of a word is its property of referring to something in a certain way, its mode of reference.¹⁶

The shocking idea: The meanings of some words, including names and natural kind words, are causal modes of reference that are partly external to the head. (Devitt 2001)

For example, the meaning of 'Mark Twain', or the property ascribed by meaning sentences involving 'Mark Twain', is the property of referring by a certain causal-historical chain where tokens of the name are borrowed from speaker to speaker, all the way to Mark Twain himself. There are other chains related to 'Samuel Clemens' that terminate in the same object, and because the chains are different, so are the meanings. Importantly, this is not to say that the meaning of 'Mark Twain' and 'Samuel Clemens' are constituted by facts that can be described in purely causal terms, for both borrowing and dubbing are forms of intentional action. The crucial difference, in which intentionality figures in the causal-historical picture as opposed to the truth conditional/descriptivist picture, is that the former appeals only to

¹⁶ According to Devitt, a token expression can have many meanings, hence the indefinite article. The issue is not relevant for present concerns, however.

particular intentions of particular speakers at particular times. There might be nothing qualitatively shared by all the referential intentions in the chain forming the meaning of 'Mark Twain' other than that they are all causally related to each other in various ways. In particular, there need not be any single description or even a cluster of descriptions, which each speaker in the chain would have to know or believe about Mark Twain in order to become a link. Furthermore, the difference in meaning construed as a difference in causal-historical chains explains why someone might be surprised to discover that Mark Twain is, in fact, Samuel Clemens and vice versa, and that the names are, in fact, coreferential. The surprise is possible because it is often surprising to come to know, especially with long-standing and widespread names, that they descend from the same origin.

If these points are even broadly correct, then the meaning of 'Samuel Clemens' is empirical. It is constituted by certain empirically observable events in the world, namely chains of speakers borrowing the name from one another. It is clear, of course, that I am using 'empirical' in a somewhat promiscuous sense, counting both historical and intentional facts as something that can be observed. Perhaps the best way to think of 'empirical' in this context is by the contrast to (a) the mental realm understood as the inner world of qualia, sensations, associations, etc., and (b) the abstract realm inhabited by entities, such as propositions and sets of definite descriptions. Indeed, Devitt's purpose in calling his idea 'shocking' is that it effectively dismantles these two traditional construals of the realm that meanings are thought to inhabit: he covers both under 'the Cartesian dogma', which understands a fundamental feature of meanings to be that they are transparent and immediately known to the subject who uses them (2001, 479).

The tricky issue for Devitt, however, is how to understand a second batch of fundamental questions about semantic descriptions. If the meaning of 'Samuel Clemens' is a certain causal-historical chain, how is knowledge about this chain 'packed' into the word so that it can be conveyed in a speech act? How is it that the meaning sentence 'Samuel Clemens' means Mark Twain' *describes* the name 'Samuel Clemens'? Devitt's general answer here is that linguistic competence should be considered on the model of knowledge-how instead of knowledge-that; that at stake is a kind of a skill and not knowledge of, say, propositions or some mental entities (2001, 481).

My main point here is that Devitt's shocking idea, if defensible, offers an interesting middle ground between MD_{LB} and the sceptical solution's deflationism concerning the nature of semantic descriptions. On the one hand, at least some semantic descriptions have substantial content according to the shocking idea in the sense that their truth involves more than the disquotational scheme. Meaning sentences can have truth makers. On the other hand, these truth makers do not have to meet the impossible standard of absolute determinacy, for as empirical facts, they are as temporal as anything.

What about expression types that do not fit so easily (as if the case of proper names and natural kinds would be easy!) the causal-historical picture, such as functional expressions, terms denoting social kinds, fictional names, etc.? I have no hope of entering that vast debate in this context; to repeat, my modestly progressive aim has been to argue that the causal-historical account can explain, in programmatic terms, how meaning sentences can work as substantial descriptions while rejecting absolute determinacy of meaning. In order to continue developing the program, in the next section, I shall show, with Mario Gómez-Torrente, how the causal-historical account can explain reference-fixing without presupposing descriptivism in characterising referential intentions.

7.3.3 Specifying Referential Intentions

From Section 7.3.1 above, we recall that Kusch's main reason for rejecting the causal-historical account as a solution to the challenge is that, according to him, the reference-fixing intentions of proper names and natural kind terms must include some descriptive elements that are true of the referent in order for the fixing to be determinate. But any such description is liable to be gerrymandered by alternative sceptical hypotheses of the type *mountain/quontain*; hence, reference-fixing intentions will always be indeterminate.

Although Mario Gómez-Torrente (2019) does not align his development of the causal-historical picture in terms of the sceptical challenge, I believe that his proposal can work as an outline of a solution, which, if successful, will counter the *qua*-problem of reference-fixing and, I argue, answer Kusch's sceptical rejection of the

causal-historical account. His proposal, as I understand it, involves two key elements: referential conventions and a metaphysical distinction between ordinary and scientific kinds.

The *qua*-problem, as Gómez-Torrente reads it, is essentially a problem of individuation. Numerous descriptivists argue that causal relations alone cannot individuate the reference of natural kind terms simply because there is considerable indeterminacy present at the level of the sample tokens, which cannot be resolved without reliance on individuation by mental descriptions, i.e. a descriptive element in the reference-fixing intentions. The indeterminacy applies both 'upwards' and 'downwards' from the samples; it is indeterminate which higher class is individuated by the introduction (hence transmission) of a term and which class of samples mixed in with the one intuitively referred to is being identified. Even if the causal mechanism is conceptually without contradiction, its truth would entail massive reference failure for our ordinary use of natural kind terms, which is not acceptable (Gómez-Torrente 2019, 159).

Gómez-Torrente's strategy is to accept the indeterminacy of our ordinary natural kind terms, yet deny that this entails massive reference failure and hence that descriptive mental contents would be required.¹⁷ What makes this possible is that, first, the referential conventions, which govern the assignments of referential failure and success among ordinary speakers, are vague, flexible, and only roughly sufficient for determining reference. An example of the several conventions he names:

Successful explicit natural kind noun introduction via perceptual intention.

If a speaker S forms the explicit intention of using a noun N that he or she introduces to refer to the natural substance (/species/phenomenon) exemplified by (most of) a bunch of items that he or she has clearly perceived, and most of the items in question do exemplify a single substance (/species/phenomenon) A, then N as used by S will refer to A, if S forms no intention conflicting with that intention, and if S doesn't form intentions about how to use N that on the whole conflict with the intentions of the community of users of N at large. (2019, 160)

¹⁷ In fact, he argues, quite ingenuously, against descriptivism by claiming that it cannot account for independent data on the referential indeterminacy of proper names (2019, 73). In other words, descriptivism fails because its predictions about determination of reference are too strong in relation to intuitions. Likewise, Kripkenstein's basic point against Meaning Determinism is that it makes too strong, deterministic assumptions about meaning's extensional dimension.

There are a lot of words in this articulation of the convention that would require attention to make explicit, but the most important here are 'exemplification', 'substance/species/phenomenon', and 'intention'. The key point is that in Gómez-Torrente's rendering, these words have what we might call minimally determinate content in virtue of their ordinary meanings, such that indeterminacy rarely arises for speakers introducing natural kind terms into language. To begin with, take the impurities or downwards *qua*-problem:

[I]t is arguably a consequence of the ordinary meaning of 'exemplification' (or of 'being a case' or 'being an example') and of the contents or *a priori* connections that particular substance nouns come to have (in virtue of being substance nouns) that the items typically thought of as exemplifying a particular substance will not also exemplify its impurities. For example, the (alleged) instances of water [...] available to ordinary speakers are not things that, given the ordinary meaning of 'exemplify,' exemplify also the kinds calcium salt, magnesium salt, sulfate, nitrate, silica, iron, sodium, aluminum, etc. (2019, 166)

A similar treatment works for the upwards side of the *qua*-problem:

While it is undoubtedly true that an instance of a given natural kind will at the same time be an instance of 'higher-level' natural kinds, these kinds will in general be implied not to be substances, species, or phenomena by appropriate meaning-constituting or *a priori* principles. For example, though instances of water are also instances of the kind *liquid*, *liquid* is not a substance, as it is compatible with the meanings of 'substance' and 'liquid' that different liquids may be different substances—and in fact, there are liquids with very different natures, which in no sense can be said to belong to one same substance. Though instances of gold are also instances of the kind *metal*, *metal* is not a substance, as it is compatible with the meanings of 'substance' and 'metal' that different metals may be different substances. Though particular tigers are also instances of the kinds *feline* and *mammal*, *feline* and *mammal* are not species, as it is compatible with the meanings of 'species,' 'feline,' and 'mammal' that different felines and different mammals may be of different species. (2019, 167)

There are exceptions this strategy must deal with individually, but overall, it should be clear how the first element of the solution is supposed to work. The intentions, along with the words 'exemplification' and 'species/substance/phenomena' included in the referential conventions, have minimally determinate contents in that they do not specify (describe) the superficial characteristic of the referred-to kind. Yet, the weak determinacy is enough to *a priori* settle, in the ordinary meaning of 'exemplification,' that what is exemplified when pointing to a puddle of water is not the mud-particles but the *water*. Similarly, for higher-class indeterminacy, which does

not arise because the ordinary meaning of, e.g. 'substance', *a priori* forecloses the possibility that a higher class, say *liquid*, is what is being referred to, for *liquid* is not a substance.

What might make this explanation unsatisfactory in the eyes of a critical descriptivist is the notion that the Kripke-Putnam orthodoxy is assumed to explain how ordinary kind terms refer to the hidden *essences* of natural kinds, by which are usually understood as jointly necessary and sufficient conditions. The second element in Gómez-Torrente's strategy is to distinguish between the so-called ordinary kinds and precise scientific kinds, both of which are rightly called 'natural'. He then argues that although ordinary kind terms are *a priori* loaded with an implication of a 'hidden essence', it is not similarly *a priori* that this hidden essence is identical with the jointly necessary and sufficient conditions, which philosophy uses in discussing essences:

It is unclear that it is part of the ordinary meaning of 'species' that a species, such as the conjectured species that the tiger is, ought to be *reducible* to scientific conditions detailing underlying structures. For it would seem that it is compatible with the meaning of 'species' and 'tiger' that the tigers might have a hidden underlying structure and yet that this hidden underlying structure might not be enough to characterize tigers. Anyone tempted by the contrary view ought to recall that it was not so long ago that it was generally thought that living beings had a crucial characteristic (life) which was not reducible to scientific conditions, let alone to conditions specifying hidden underlying structures. Presumably vitalists did not have in mind non-ordinary meanings for 'species' and 'tiger.' Surely, similar views about the nature of water were possible (and probably actual, especially among people for whom water had religious purifying powers of some sort). (Gómez-Torrente 2019, 171)

What this entails is a significant split between ordinary kinds as referred to by ordinary natural kind terms, and precise scientific kinds as referred to by scientific terms:

When one thinks carefully about it, it actually appears very reasonable to conclude that the conjectured ordinary kinds *water, gold, tiger*, etc. are, if anything, kinds not reducible to scientifically precise necessary and sufficient conditions given in terms of underlying structures. One reason is in fact provided by a gestalt switch, as it were: we can see arbitrariness arguments not as eliminativist arguments showing that no reference is fixed because the choice among scientifically precise structural kinds is arbitrary, but as revealing intuitive constraints on the kinds possibly referred to. And once one sees this, the constraints in question point toward the view that the kinds referred to, if they exist, are in fact kinds irreducible to scientifically precise structural kinds. (Gómez-Torrente 2019, 173)

So, what are 'ordinary' kinds then, if not precise, scientific kinds? Are they really 'real' or not? The short answer Gómez-Torrente gives is that there is no reason to think ordinary kinds as any less 'real' than scientific kinds since we cannot reduce the former to the latter. Why they cannot be reduced is partly because of differences in modal properties, as argued by Kripke in unpublished lectures and Mark Johnston (1997), but also because they have different 'indeterminacy profiles': for example, it is not determinate whether the ordinary kind *water* has samples, which are also samples of 'heavy water' (D₂O), while it is determinate that H₂O has no such samples (Gómez-Torrente 2019, 176).

So far, so good. But how is all this to be squared with the sceptical challenge? Namely, what to make of Gómez-Torrente's frequent appeal to 'ordinary meanings' of terms, such as 'exemplification', which he claims involve constitutive *a priori* relations to other key terms. Where these are used to counter the *qua*-problem, which Kusch correctly identifies as the wedge into gerrymandering, the sceptic can ask: what is the fact that determines that, e.g. 'exemplification', has the *a priori* meaning Gómez-Torrente claims it has? Perhaps what is actually meant is 'quexemplification,' where... But we can already stop there.

My suggestion is that, instead of reading Gómez-Torrente's invocation of ordinary meanings with *a priori* content as allegiance to Low-brow Meaning Determinism, we instead interpret this as an affinity to the sceptical solution's deflated understanding of meaning ascriptions. Indeed, an important aspect of his strategy is to construe referential intentions as something devoid of descriptive content, which is precisely contrary to how the MD_{LB} construes intentions to use '+'. Most of the work for individuating reference is done by the *conventions* together with the causal facts of the matter, minimising the appeal to descriptive contents. The flexibility, defeasibility, and indeterminacy with which Gómez-Torrente characterises the conventions also very much resembles the rough-and-ready assertability conditions of the sceptical solution. Moreover, Gómez-Torrente does not, in the book, submit to any specific general theory of conventions, which he argues are already implied as a part of the solution by Kripke in *Naming and Necessity* (2019, vii). In other words, the causal-historical theory of reference does not need to challenge so much as complement the sceptical solution, where it is stripped free of the remnants of Meaning

Determinism lingering in descriptivism forced upon it by a wrong resolution of the *qua*-problem.

One mystery remains however: why does not Kripke, the author of the causal-historical theory of reference, not propose it as a solution to the challenge if it squares so well with the sceptical solution as I claim? Kusch wonders the same thing, one reason for which might be that the mathematical example is not a strong point of the causal-historical theory of reference. But Kusch also mentions a stronger, more condemning reason, which is that the causal-historical theory of reference 'is unworkable as an answer to the sceptical argument' (2006, 136).

That is condemning indeed. However, it should be observed that the reason Kusch has for the judgement draws from developments of the causal-historical theory of reference that are not signed by Kripke himself, namely those due to Colin McGinn (1984) and Penelope Maddy (1984). It is questionable whether Kripke would have opted for defences, which rely on neural reductionism about mental states, for example. Furthermore, there is a reason Kripke gives in WRPL that appears to explain the omission of the causal theory already in the Preface: 'It deserves emphasis that I do not in this piece of writing attempt to speak for myself, or, except in occasional and minor asides, to say anything about my own views on the substantive issues' (ix). Since the causal theory of reference most certainly belongs to the category of Kripke's own views, the mysterious omission dissolves naturally.¹⁸

However, although Kripke's omission of the causal theory of reference in WRPL can be read neutrally concerning his views about its plausibility as an answer to the challenge, this does not entail that the causal-historical theory and the sceptical solution make uncomplicated bedfellows. For one, we cannot say for certain what Wittgenstein would have thought about the matter. But insofar as we do not mix substantial issues with exegetical ones, what is there in the sceptical challenge and solution, as read by Kripke and Kusch, that is incompatible with the causal-historical account of reference, as read by Gómez-Torrente?

¹⁸ Relating to Kripke's own expressed views in WRPL (66), he also remarks that: 'Personally, I can only report that, in spite of Wittgenstein's assurances, the 'primitive' interpretation often sounds rather good to me...'. I confess not to know what to make of that.

To offer a convincing, full answer to that question would be more than I can provide in this work. Nonetheless, I hold to the idea that the connection between the sceptical challenge and solution and the causal-historical theory of reference should not be so quickly dismissed as Kusch does. What is truly problematic for the sceptic is Low-brow Meaning Determinism, which is smuggled into the causal-historical theory in the way of descriptivism, and Gómez-Torrente makes powerful arguments why descriptivism in all its shapes is repelled by the causal-historical theory of reference. If that is true, the question of compatibility boils down to how closely referential conventions resemble sceptical minimal factualism. Above, I showed that there are at the very least *prima facie* resemblances in:

- a) *Flexibility*: Both referential conventions and assertability conditions can change due to contingent social causes.
- b) *Vagueness*: What the conventions and assertability conditions exactly are and how they should be articulated is vague.
- c) *Defeasibility*: Neither conventions nor assertability conditions articulate the jointly necessary and sufficient conditions of individuation, but only roughly sufficient ones.
- d) *Publicity*: While neither assertability conditions nor referential conventions are usually explicitly known by the community, they are essentially *shared* by the subjects in some sense.

There's no smoking gun here to be sure (a slightly bloody knife at most), but that does not mean these resemblances are not genuine clues to something more substantial.

7.3.4 Aftermath of the Challenge: Living with Indeterminacy

When all is said and done, has the sceptical solution been solved by this chapter? The answer can be nothing else than yes and no. The main idea that I have pressed in this chapter is that the first step to solving the sceptical challenge must be to get clear about the real problem, which, I have argued, is finitude rather than normativity. In particular, in its primary form, the problem of finitude is logical rather than metaphysical, the main pivot being the notion that in order for any actual

application of an expression to be determinately correct or incorrect, all logically possible applications of it must be. This is, roughly, the core idea of absolute determinacy. Only once we give up absolute determinacy and the atemporal picture of meaning determination, which it implies, does it become possible (or at any rate, useful) to think of the challenge as a metaphysical problem.

The second overarching idea that I have argued for in the chapter, albeit briefly, is that the causal-historical account of reference promises to be a fruitful way to develop the metaphysics of meaning in the temporal fashion promoted by the aftermath of the challenge. Indeed, the thought that the determination of meaning is a historical process, in a sense, inbuilt in the account, especially in the later developments that include multiple groundings (Devitt 1996, 167), partial references (Devitt 1996, 228), and referential conventions. Moreover, rejecting absolute determinacy as impossible for finite speakers to achieve allows a new view on the notorious *qua*-problems. What is crucial here is that a successful solution to the *qua*-problems need not appeal to speaker-intentions, understood as descriptions of the referent that would have to be absolutely determined; instead, it is accepted that there will always be a residuum of indeterminacy in reference-fixing, at least when dealing with macro-physical noun-terms.

It goes with the causal-historical account, as I understand it, that the residuum of indeterminacy is a feature and not a bug of natural language. Not that there is nothing to be done about it. Indeed, scientific language (as any specialist language) is arguably an effort to reduce the indeterminacy of ordinary language.

CHAPTER II: ALL GENUINE NORMS GO TO TRANSCENDENCE

8 PLAN FOR THE CHAPTER

This chapter is meant, first, to introduce the core architecture of Brandom's work on meaning, language, and discursive norms. In particular, I shall focus on his efforts to synthesise certain realist and anti-realist claims about meaning. At the heart of his thinking, I believe, is the combined drive to conserve the traditional commitment to propositional contents as central to understanding linguistic meaning, while also taking seriously the fundamental pragmatist criticism of such contents where they are understood to exist independently of human social practices. In Brandom's terminology to be elaborated below, the synthesis he is after sparks between the commitments to *phenomenalism about norms* and *normative phenomenalism*.

Even as Brandom is engaged in the Herculean task of renovating traditional analytic philosophy by a paintwork of classic pragmatist ('classic' in the sense of Kantian and Hegelian) tones, he is keen to contrapose his 'analytic pragmatism' with the more recent externalist and representationalist order of explanation, as most famously represented by the Kripkean legacy and the new work on a theory of reference. The key objection that Brandom levies not only against Kripke's theory of reference, but also the reliabilist theory of justification is that merely causal relations and processes are insufficient to explain the salient features of the intentional nexus, such as meaning, reasons, justification, representation, etc. What is also required is the *genuinely* normative element, by which Brandom means discursive norms implicitly instituted by practices.

The second task of the chapter is to pose criticism to the very idea of implicitly instituted normativity where it stands for an exclusive alternative to the 'causalist picture' that was developed at the end of Chapter 1. In particular, I am interested in challenging Brandom's contention that the mechanism by which meanings are determined must involve genuinely normative *justification*. The focus of the criticism will then gradually home in on the theory's pragmatic foundations – most

importantly, the nature of genuinely normative *force* and how it is said to be 'instituted' by the practices.

As lengthy as this chapter admittedly is, it is forced to remain incomplete. In Chapter 3, we will focus on Brandom's recent work, called *A Spirit of Trust* (2019), which is a significant part of this discussion.

9 THE CORE ARCHITECTURE

9.1 Brandom and Kripkenstein

Brandom's relation to Kripkenstein's sceptical challenge is complicated. On the one hand, Brandom overtly approves of the criticism against the dispositionalist theories of meaning (MIE, 29), and his pragmatist-inferentialist theory fits well with the rejection of absolute determinacy. On the other hand, Brandom criticises Kripke specifically for a too narrow reading of later Wittgenstein's thought on meaning and specifically its relation to normativity. In a nutshell, Brandom and Kripkenstein agree it is a mistake to view meaning as a static state or a fixed property (in the sense of absolute determinacy), for it is rather an evolving, complex process, the roots of which are in subjects' practical engagement with each other and the world. At the same time, Brandom sees himself as diverging from Kripke's reading of Wittgenstein in claiming that the process is essentially normative in nature, while further diverging from (his reading of) later Wittgenstein himself in that the process can be given a coherent, encompassive philosophical explanation. So, while meaning is a normative process, normativity is not a primitive phenomenon for Brandom.

This is where I think my reading of Kripkenstein, which follows that of Kusch, most starkly diverges from Brandom's landscape. For in my view, Kripkenstein convincingly shows that there is no genuine normativity, in any sense of the term, implicated in meaning something by a word. In particular, the mechanism by which meanings are indeed determined does not hinge on any kind of metasemantic justification. My full argumentation on this point can only be made to bear in Chapter III, however.

How do Brandom and Kripkenstein agree on what meaning is *not*? We first encounter Brandom's rejection of the absolute determinacy of meaning in his claim that 'objectivity is a structural aspect of the social-perspectival *form* of conceptual contents' (MIE, 597). What this means is that the determination of discursive

commitments that are truly correct, i.e. which assertions and inferences should be endorsed (in context), can only be settled from individual doxastic perspectives, where it appears as the difference between two sets of commitments – those which the subject herself undertakes (which she takes to be true), and those which she attributes as undertaken by others. The first perspective 'keeps the book' on the commitments that others ought to acknowledge (according to the scorekeeper), whereas the second book tracks those commitments they, in fact, are disposed to acknowledge (again, according to the scorekeeper). This 'scorekeeping idiom' will come into a more detailed focus in Section 9.2. Brandom's key claim is that propositional, conceptual contents are objective precisely *because* the same contents can be specified from two different doxastic perspectives by every scorekeeper in the practice; the contents' objectivity *consists in* their essential social-perspectivity.

It is crucial to get this gestalt shift right, in an outline initially, for it is at the heart of the rethinking of propositional contents that Kripkenstein's sceptical challenge started and which Brandom's theory independently pursues. The received, criticised view on propositional contents comes from Frege, for whom propositions were mind-independent, abstract objects that have fully determinate truth conditions (concerns of vagueness notwithstanding). The emphasis on problems for Frege is not on the metaphysical issue of how absolute determinacy is possible to begin with, but on how finite subjects can have epistemic access to or 'grasp' the contents. How can the subjects form a sufficiently encompassing subjective perspective on the objective contents in order to count as relating to them in judgements, assertions, etc.?

In contrast, Brandom's initial question is, not how the subjects can grasp the contents, but how the contents (norms) can get a grasp on the subjects. The key pragmatist-phenomenalist methodological constraint that Brandom vests on propositional contents is that they must be understood as the products of normative social-perspectival practices (MIE, 25). The primary sense in which the contents (i.e. norms) are to be conceived as 'objective' is that by the scorekeepers' 'own lights', every reason asked and given can be found incorrect from another, perhaps never instantiated, doxastic perspective. In order for that to be possible, *phenomenalism about norms* must be balanced by *normative phenomenalism*, where what is correct, i.e. which propositional contents are actually *conferred* on expressions, is determined not by the

attitudes, but by the norms *instituted* by the attitudes. In MIE, 'institution' means precisely that discursive perspectivalism constitutes the objectivity of contents as the form in which the contents *appear* to the subjects (MIE, 584).¹⁹

It is also important not to confuse formal 'objectivity' (or 'correctness') with substantial 'truth'. A parallel social-perspectival account about conceptual contents that Brandom rejects, dubbed *I-we* sociality as opposed to his *I-thou* variety, holds an epistemic account of truth, such that whatever the community of subjects, or some select group of experts among them, would take to be correct in ideal conditions or at the 'end of inquiry' is the definition of truth. The *I-we* account postulates a privileged perspective, e.g. the whole community's or the experts', the judgements of which in at least some ideal conditions would be logically identical with truth. As a deflationist, Brandom denies the general possibility of giving a substantial definition of truth (MIE, 326), but he also abandons *I-we* sociality on the more particular grounds of the *I-thou* variety, according to which:

[t]here is only the actual practice of sorting out who has the better reason in particular cases. The social metaphysics of claim-making settles what it means for a claim to be true by settling what one is doing in *taking* it to be true. It does not settle which claims *are* true—that is, are *correctly* taken to be true. That issue is adjudicated differently from different points of view, and although these are not all of equal worth, there is no bird's-eye view above the fray of competing claims from which those that deserve to prevail can be identified, not from which even necessary and sufficient conditions for such deserts can be formulated. The status of any such principles as probative is always itself at issue in the same way as the status of any particular factual claim. (MIE, 601)

We can understand Brandom's claim here best by seeing it as a compromise resolution between two prior theses. The *I-we* sociality account claims that there are *objective propositional contents*, such that we can define their truth as whatever would be *held* true by some epistemically privileged perspective. The material negation of this claim comes in the sort of anti-realist stance championed, e.g. by Michael Dummett, who (roughly) identifies meaning with an understanding of meaning. In this view, there is no epistemically privileged perspective on truth *qua* which to objectively

¹⁹ I will not seek to give a strict definition for either phenomenalism about norms or normative phenomenalism, for I believe the spirit of these commitments in Brandom's works is primarily programmatic and thus strategically malleable according to the context, although it is also true that the fact exposes Brandom's key claims to hindering polysemy, as noted, e.g. by Jeremy Wanderer (2008, 74, fn.). The dynamic content of these commitments will unfold during the course of this chapter.

individuate meanings *because* there is no objective meaning at all, only individual subjects and whatever they can come to *understand* (take or treat) as meaningful in their mutual interactions (Dummett 1978).²⁰ According to Dummett, the perspectives cancel out both the need and possibility of objective meaning, following the Quinean idea that semantically relevant differences between words must manifest in some possible behavioural circumstances. Both the 'objectivist' *I-we* account and the 'subjectivist' Dummettian anti-realism agree (a) that meaning must be social-perspectival (every relevant semantic difference must manifest in some possible behavioural or epistemic circumstances; cf. 'linguistic behaviourism'), and (b) that if there is no epistemically privileged perspective *qua* which to define truth, then there are no objective meanings.

Brandom's synthesis accepts the (a)-premise but denies (b). In criticism of Dummett, he writes:

The good thought is the idea, cognate to that common to logical and philosophical behaviorism, that the point of invoking meanings is to explain (proprieties of) the use of linguistic expressions. But that good thought does not, as Dummett thought, preclude explanations that appeal to items not definable in terms of the linguistic behavior they are invoked to explain. That is, it does not preclude the invocation of meanings as only inferable from specifications of linguistic behaviour in a nonsemantic vocabulary, rather than definable without remainder in such a vocabulary. It does not require jettisoning the idea of truth-conditions in favor of assertibility conditions, or reference in favor of recognition conditions. There are constraints imposed by understanding truth-conditions and reference as theoretical postulates invoked to explain, for instance, the norms that practically govern fact-stating assertions of sentences and object-recognizing uses of singular terms. But those methodological norms are not so restrictive as to forbid the semantic notions appealed to inferentially outrunning what is observable at the level of non-semantically described linguistic behavior. (Brandom 2019, 17)

Brandom's reason for rejecting linguistic behaviourism is *not* that, as Kripkeans or Putnamians would have it, many concepts would have their meanings in part determined by (modal) facts about the natural kinds to which they refer, for Brandom is a deflationist about reference and of truth in the sense that neither is a kind of property. What are, then, the semantic inferential conclusions that can 'outrun' descriptions of observable linguistic behaviour in non-semantic vocabulary? The answer is delivered, of course, by *I-thou* social perspectivalism. The difference to

²⁰ For a detailed, critical examination of Dummett's semantic anti-realism, see Panu Raatikainen (2013).

the metaphysical realist doctrines of Kripke and Putnam is that the 'externalist element' is identified by Brandom as the structural feature of scorekeeping perspectives rather than as a (modal) fact out in the world.²¹ It is a structural feature of every scorekeeping perspective (which are all made equal in *this* respect) that every commitment undertaken or attributed is in principle open to be exposed as incorrect from some other perspective.

But does that essentially negative claim have any positive consequences regarding the objectivity, and hence reality, of meaning? Does semantic vocabulary, as understood in the structural I-thou social-perspectival sense, underlie any *real* phenomenon, or does it simply mark a kind of limit of our epistemic and semantic practices?

As we shall later see, this general question reveals a profound problematique at the heart of Brandom's theory. In programmatic terms, his challenge is to combine phenomenalism about norms with normative phenomenalism, or the idea that norms (hence meaning and propositional, conceptual contents) are products of our discursive attitudes with the idea that the attitudes in turn are genuinely *causally responsive* and *normatively responsible* to the norms (hence meanings and concepts) *themselves*, i.e. as something distinct from the sum of attitudes.

Before moving on to examine how Brandom may or may not plausibly pull off his grand synthesis between phenomenalism about norms and normative phenomenalism, it is useful to describe in an outline how he thinks where and how Kripkenstein goes wrong. For one, the way Brandom can solve the problem of finitude that proved fatal for dispositionalism (and to MD_{LB} in general) is rather simple. As we recall, the insurmountable problem for dispositionalism is what actual fact could determine which ascription of two incompatible meaning-determining functions to the subject is true and which is false. Because Brandom gives up the notion that there must be, or even could be, an attitude-independent truth of which function one is really following under the thesis that conceptual contents are

²¹ In fact, Brandom is a modal realist, although for different reasons than Kripke or Putnam. Furthermore, the important point here is that Brandom's commitment to modal realism is independent of his account for the objectivity of semantic contents, unlike it is for Kripke, Putnam, and other 'traditional' semantic externalists. The issue of modal, or conceptual, realism will come into focus in Section 11.5.

essentially perspectival in the I-thou sense, the gerrymandering problem does not arise for him in the same sense it does for dispositionalism.²²

A problem that arises for Brandom is why and how anyone could be *genuinely justified* in attributing one meaning-determining function as opposed to a different one to the speaker. The issue here is not so much with *what* exactly is being ascribed (i.e. the content) as to how the *act* of ascription, or more primitively that of attribution, of meaning is supposed to work if it is to be genuinely normative. Whereas Kripkenstein, according to my and Kusch's readings, claims that there are no genuine justifications available for the applications of words and attributions of meanings, or else that the 'justifications' end in brute responses, agreement, and other causal factors, Brandom states it is illicit to:

restrict one's specification of that practice of application, as Kripke implicitly does, to what can be stated in a *nonnormative* vocabulary. Using an expression *correctly* or *incorrectly* is also something we *actually* do. (TMD, 214, fn.; see also, Brandom 2001, 605)

What Brandom means is that the mechanism by which meanings are determined, and by which erroneous applications are identified, essentially involves genuinely normative justification. Although no substantial extensional notions, such as truth, reference, or representation, figure in the justification of attributions of meaning, Brandom firmly believes that *something* must justify the attributions for them to be meaningful, and that the justification must amount to something more than (a) our ordinary ways to justify ascriptions of meaning or (b) some purely externalist theory of justification like reliabilism. This is the sense in which meaning is 'essentially normative' for him – because the justification of meaning attributions is.

In order to understand Brandom on this point, it is necessary to become acquainted with his arguably most important technical and philosophical contribution – the scorekeeping practice. In the following sections (9.2 and 9.3), I shall focus on the

²² That is, of course, the short answer, and Brandom has been challenged on this point; see especially Anandi Hattiangadi (2003) and Daniel Whiting (2006). To foreshadow, in this and the following chapter, my idea is that the problem of finitude, which concerns the objectivity and determinateness of conceptual *contents*, is (for Brandom) less of a problem than usually understood in the literature. Instead, his major problems concern the objectivity and determinateness of discursive normative *force*, or the normativity problem as opposed to the finitude problem.

two most important features of the model, namely its social-perspectival and historical nature.

9.2 Meaning as Social-Perspectival

In the previous subsection, it was stated that for Brandom, meanings (propositional, conceptual contents) are essentially social-perspectival. The aim of this subsection is to explain what that means in technical detail. Unless otherwise mentioned, all the references in this subsection are to Chapter 3 of MIE.

To begin with, as a model, Meaning as Social-Perspectival what is it exactly that the scorekeeping practice is supposed to be a model about? The short, and mostly encompassive, answer is the business of making *claims* and *material inferences*. The first task is to explain these basic concepts, which only make sense when related to each other according to Brandom.

To make a claim is to produce a performance that is *propositionally contentful*. In what might be called the representationalist view, to be propositionally contentful at minimum means to have a truth value, and in a more robust sense, to have truth conditions. Not so for Brandom. Instead, to be propositionally contentful primarily means to serve as a premise and as a conclusion respectively in material inferences. To be a claim, then, implies the ability to serve as a reason for further claims, and something for which other claims can be offered as reasons for. Indeed, for all relevant purposes, a claim here just is whatever can play the dual role as a premise and as a conclusion, for I shall not venture to explain any of the vital extensions of the scorekeeping practice, e.g. how it is supposed to explain the propositional contentfulness of actions and perceptual states. Furthermore, I will not explain the central ISA structure (inferential, substitutional, anaphoric), according to which subpropositional conceptual contents can be explained all the way down to non-repeatable indexical and demonstrative expressions in substitutional-inferential terms, or why specifically propositional contents are the paradigm form of conceptual contents.

At this point, it might be interjected that an account of what it is for discursive subjects (or 'scorekeepers') to take or treat certain performances as propositionally contentful, respectable as such an account might be, should obviously be distinguished from what such contents are in themselves. That is, we should firmly distinguish the pragmatics of what it means to treat performances as *claimings* and the semantics of what is thereby *claimed*. Alternatively, since Brandom's theory can also be understood in the tradition of expressivism, we should distinguish between asking what it is to express propositional contents and what propositional contents are 'in themselves'.

It just so happens that the scorekeeping practice is sensitive to this difference, and more, is posed to explain it as well. In particular, the core architecture of MIE is to explain propositional contents (or what is claimed) by the acts of claiming (and in particular, asserting) them. This kernel of 'fundamental pragmatism' further forks into two distinct pragmatist theses, namely methodological and semantic.

On the methodological side, the claim is that what the postulations of propositional contents (meanings) are *for* is to explain proprieties of the practice of making claims and inferences. Although it makes sense in the specific case of formal languages to posit meanings for symbols without worrying what greater explanatory purpose their postulation serves, in case of a real phenomenon (such as natural language), the postulation of theoretical terms (such as meanings) should always in principle serve an empirical explanatory end. (Remember that Brandom differs from Dummett and Quine on this point by denying that it would therefore be illicit to postulate any meanings that would not be manifested in any linguistic, or more broadly epistemic, behaviour.)

On the semantic side, the pragmatist claim that Brandom pursues is the foundationalist thesis: what ultimately glues together a given meaning postulated by the theorist with the expressions of the natives is the *use* made of the expressions. The use is not describable in terms of a special class of facts – say, semantic

dispositions of the practitioners, but in terms of specifically normative vocabulary – that is to say, in the vocabulary of scorekeeping, to which we must now turn.²³

The most basic kinematics of the scorekeeping practice can be understood as the interactions between two different normative *attitudes* and two different normative *statuses*. On the side of attitudes, we have *attribution* and *undertaking*, while on the side of statuses, we have *commitments* and *entitlements*. The three basic interactions in the practice consist of attributing a commitment and/or entitlement to another subject, or then undertaking a commitment oneself. Of these, the most important move is attributing a commitment to another subject because undertaking (or acknowledging, which is a self-aware, explicit form of undertaking like ascribing is an explicit form of attributing) a commitment just is to do something, entitling the audience to attribute the commitment.

An illustration will be helpful at this point. In Aesop's fable *The Boy Who Cried 'Wolf!*, a bored shepherd lad finds the repeated abuse of his power to summon nearby villagers in aid against an imaginary canine predator, thwarted by the said villagers' withdrawal from him the precise normative franchise. What this means in scorekeeping vocabulary is that whereas they had previously attributed to the subject both the commitment *and* entitlement to the claim expressed by his cries that there is a dangerous canine beast nearabouts, after perceiving repeated failures to live up to the responsibilities involved in the observational commitment, the villagers cease treating the boy's commitments as entitled. They *negatively sanction* his performance, as they no longer treat his calls for help as authoritative to oblige their input. The sanctioning can further be understood either in an *internal* or *external* sense. In the internal sense, which alone is genuinely normative, the villagers treat the subject as unauthorised to leverage their trust in him; although they are authorised to attribute the responsibility over the veridical observation of canine predators to him, they do not themselves undertake the purported deliverances of the observations.²⁴ In the external sense of negative sanctioning, the villagers causally affect the subject's ability

²³ In MIE, Brandom does not explicitly distinguish between methodological and semantic pragmatism, but he has later recognised that though they make natural allies, the two theses may also be pursued independently from each other (PP, 63, fn.).

²⁴ In a more radical sense, the villagers might also deny the boy's cries the power to assume responsibility itself over veridical observation, effectively ceasing to treat his voicings as meaningful at all as opposed to false alarms.

to incite further performances on their part by affecting his dispositions to commit beckoning performances, in this instance, by leaving him for the wolf to sort out, or perhaps in some kinder, less karmic world by beating him with sticks instead.

It will be helpful to briefly reflect on the moral of this little fable. What Brandom claims is that certain dispositional abilities (discussed further in Section 9.4 below) are *necessary* for subjects to treat performances as propositionally contentful. He is even willing to admit that in some derivative, secondary sense, the dispositions are also *sufficient* for the task in that non-human animals can also be treated as having beliefs and desires expressed by their performances displaying practical intentionality (discussed further in Section 12.6). However, it is crucial to notice that most of the explanatory work will ultimately fall not on dispositions but on the norms, said to be *implicitly instituted* by the villagers' internal practices of sanctioning, which, together with their external dispositions to sanction, are necessary and sufficient to confer propositional contents on their performances in the primary sense.²⁵ The idea of implicit normativity as the cornerstone of scorekeeping practice will be in focus through this chapter, most notably from Section 9.5 onwards.

Brandom's next move is to construe the scorekeeping practice proper on a specific set of internal sanctions by which the subjects track each other's commitments and entitlements, somewhat like goals are tracked in games. But what is it exactly that the villagers track when they attribute to the shepherd boy the observational commitment to the claim that 'There is a wolf nearby about to eat all the sheep'? They track its *inferential significance* – in two dimensions, no less.

First, to track the inferential significance of a commitment to a claim in the *interpersonal* sense means tracking, not only the claim the attributor takes the subject to have undertaken, but also the claims that she takes to be the claim's *commitment-preserving* consequences. For example, 'Thunder will be heard soon' is a commitment-preserving consequence of the claim 'Lightning struck over there'. This is a *material inferential* relation as opposed to a formal one, because the correctness of the

²⁵ Strictly speaking, it is only the sufficiency claim that Brandom sets as a criterion of adequacy for himself in MIE (xx), although he also claims that the account can fulfil the necessity claim (MIE, 159). However, as will turn out in Section 12, I think the necessity claim is, in fact, not merely optional to Brandom's account due to his emphasis on genuine discursive normativity as central for rationality.

inference depends on the conceptual contents of the premise and conclusion. In attributing the claim and the material inferential consequences, the scorekeeper takes it to have to the subject, the scorekeeper treats the subject as both disposed and obliged (to be prepared to) to assert those claims, and furthermore, as obligated to justify them if appropriately challenged to. The justification itself may occur either in *intracontent* fashion, as showing the claim to be a material inferential consequence of commitments, the subject is entitled to (perhaps via observation), or then intrapersonally, as deferring to another subject as the original author of the claim.

Already, we can see here what a neat fit the scorekeeping practice makes with inferential semantics, where the content of a claim is construed as the network of material inferential relations it participates in as a premise and a conclusion. An important addition Brandom adds to the tradition of inferentialism is the status of entitlement: scorekeepers differentiate between those material consequences of a claim to which the subject is committed to, and those to which he is entitled as well, or which commitment-preserving inferential relations are also entitlement-preserving. Which commitments will then be treated as entitled? In the basic sense, only those which the scorekeeper undertakes herself, taking us to the *intrapersonal* dimension of inferential significance. This is the dimension in which each scorekeeper traces the *actual* inferential significance of various claims, contrasting them with what she takes others to *take* those significances to be.

9.3 Meaning as a Historical Process

We have now seen that the thesis 'meaning is social-perspectival' means that it is, in fact, doubly perspectival. Not only is the score kept *for* each scorekeeper separately, so that no two subjects can be strictly speaking said to share all the same beliefs and meanings (propositional contents), but it is also kept *by* each scorekeeper separately. In this section, I shall further show that the intra- and intersubjective, synchronic scorekeeping practice, once understood properly as a historical and a diachronic process, is really *triply* perspectival.

To do that, I shall briefly describe a central aspect of how Brandom understands Hegel to have improved on Kant's theory of judgement. The result will be an outline

of the process in which conceptual contents are determined, called by Brandom the 'three judges' model'. Furthermore, we shall see that the Hegelian step also signals a broader transformation of what Brandom calls our 'metaconceptual' views. (As a general side-note, in this work, I will not challenge or evaluate Brandom's exegetical readings of the classics.)

According to Brandom, Kant's theory of judgement can be read as a two-step process in which, first, a determinate, *explicit* conceptual norm (a 'rule' in Brandom's vocabulary) is *selected* and, second, *applied* in a particular judgement. There are two important features in this (extremely stripped) reading of Kant's account that Brandom thinks Hegel improves on. First of all, for Kant, the rule that in the first step is selected as a candidate for the application is '*complete*' or '*maximally determinate*' because it is fully settled in advance, for every logically possible instance of the application, whether the application would be correct or incorrect.²⁶ Second, the application itself does not affect the completeness of the rule that determines the content of the concept; the two steps are operationally fully independent in their logically fixed order. If the selected rule is found to lead to errors in the second step, the process abandons it and returns to select another rule for a separate 'fitting' (TMD, 213; RP, 81).

Instead of the two-step process developed by Kant, Hegel favours a one-step process in which:

[c]onceptual content arises out of the process of applying concepts—the *determinate* content of concepts is unintelligible apart from the *determination* of that content, the process of determining it. Concepts are not fixed or static items. Their content is altered by every particular case in which they are applied or not applied in experience. At every stage, experience does presuppose the prior availability of concepts to be applied in judgement, and at every stage the content of those concepts derives from their role in experience. (TMD, 215)

Brandom proposes that the key to understanding the process of applying determinate concepts as also simultaneously determining their contents is Hegel's *reciprocal recognition*. The full exposition of Brandom's thought here must wait until

²⁶ This is, in fact, very close to the idea of absolute determinacy discussed in the previous chapter: a connection that will come up later.

Chapter 3 and the discussion of *A Spirit of Trust*; for now, a brief explication of the idea by analogy to the common law tradition will suffice.

Brandom's premise, following Hegel's one-step theory of judgement, is that maximally determinate (or absolutely determinate) conceptual, propositional contents do not exist. It is not determined in advance and in abstract for at least any empirical concept what would (not) count as its correct or incorrect application in particular cases. All there is to determine the correctness of any application is the appeal to other applications of the concept, in two respects. First, the past applications of a concept *C* exercise authority over its present applications, in analogy to how past decisions in the common law tradition exercise authority (constrain and guide) the judgements of a judge tasked to give a verdict in a present novel case. However, it is up to the present judge to decide which cases (previous applications of *C*) she finds to be both relevant and correct to justify her verdict; the past applications do not exercise any constraints *by themselves*, without mediation by the judge's discretion. Second, once the present judge makes her verdict and applies *C* to a particular case, justifying her use in terms of the preceding applications she finds as both relevant and correct, her application in turn will be exposed to similar evaluation by future judges, who then exercise their own authority to deem whether the present judge's application is correct in a context deemed as relevant. In this chain of applications that includes evaluations of correctness of preceding applications, nowhere is the point reached where the complete, maximal/absolute determinacy of *C* is reached, but the concept's content remains open both to the future *and* the past, in the sense that any past or present application may be found incorrect in the future (including evaluations of the correctness of past applications) (TMD, 230-231; RP, 84-87).

Already, this brief exposition of the three judges' model raises sharp, troubling questions for Brandom. Seen from the perspective of a present judge investigating the past applications of a concept, the task is naturally construed as *epistemic* in character, i.e. a matter of *finding* the correct interpretation among the relevant set of cases. Whatever presently is the correct application is, in principle, already implicit in the past applications, to which the judge's discretion must be responsive if she is to make her judgement rationally justified. In contrast, when we look downstream from the present judge's decision to apply the concept in a certain way, it will appear

that she effectively *created* a new meaning for the concept by the act of applying it in the way she did, and that her decision actively purports to oblige the future judges to recognise her decision as authoritative among all the other options available in the context. For although her decision, if sincere, will appear to her as the best and possibly the only necessary course of action, it will only rarely (or ever) appear as the absolutely singular possibility of how the concept could have been applied in the case. In any case, no matter how necessary the actual application *appears* to the present judge, nothing yet guarantees that it *was* necessary.

Now, clearly enough, there is a tension between being aware of both these perspectives immediately. The concept's content will appear on the one hand as found and determined and on the other as made and indeterminate; so, which is it really?

Here, Brandom's response follows the method of proliferating perspectives by claiming that the choice of seeing the two temporal perspectives (retrospective and prospective ones) as incompatible with each other is indeed optional and wrong, a sign that one is operating with a *Verstand* metatheory of concepts. This effectively means holding onto maximal or absolute determinacy, or the idea that a concept's content must be either atemporally fully determinate for each logically possible case of application or else the content must be indeterminate. *Contra* Kantian or Fregean absolute determinacy, Brandom proposes the Hegelian *Vernunft* metatheory of concepts, according to which the temporal perspectives are not incompatible. There is such a thing as the real, determinate meaning of a concept, but this determinacy is temporal in nature, such that whatever the real meaning of a concept is can *change retroactively in time*.

Let us pause for a moment to wonder what exactly is implied by that statement. For one, how could the *change* of virtually *anything* be possible *retroactively* in time? Currently, various philosophers and physicists, particularly those of a certain eternalist persuasion, find it hard enough to buy that the *present* or the *future* could genuinely change, let alone the past! Indeed, as an insightful commentator of Brandom, Ronald Loeffler has observed it would be reasonable to read Brandom's proposal for the *Vernunft* metatheory as effectively belonging to the meaning eliminativist tradition of Quine and Davidson (2017, 175-176). After all, since the

retroactive determinacy of conceptual contents turns out to be merely an appearance when conjoined with the prospective perspective, the conclusion might as well be that the content really is indeterminate, or determinate only in relation to a scorekeeper and her temporal perspective.

Loeffler's proposal is, I think, reasonable and broadly in line with his (2005) suggestion that the scorekeeping account is best read along meaning eliminativist lines. In programmatic terms, Loeffler's idea is that Brandom would benefit from easing on his commitment to normative phenomenalism in favour of strengthening phenomenalism about norms. Since normative phenomenalism stands for the claim that attitudes are normatively subject to the norms themselves as to their correctness, its pull is naturally towards a realist stance on meaning, whereas phenomenalism draws towards eliminativism (or reductionism) and the dominance of attitudes.²⁷

However reasonable it might be to choose between the prospective and retroactive perspectives, or phenomenalism about norms and normative phenomenalism, Brandom apparently has yet to give up a strong commitment to either. In my view, that is as it should be, for we have not yet taken a go at this latest attempt to synthesise these commitments in ST. For now, it suffices to say that, because of reasons to be later explicated, the only way Brandom can hold on to his bold synthesis is by claiming that meanings can change retroactively in time. For the moment, I will give the idea the benefit of doubt, although for the record, I will say that ultimately, I do not think the claim can be sustained.

I think that, at this point of my exposition, the most charitable way to make Brandom's radical proposal intelligible is by rethinking the concepts of descriptive and evaluative terms. It has been usually thought that while it is the business of terms like 'dog' to describe objects, it is the privilege of another class of terms, e.g. that of 'good', to evaluate them in terms of some standard. So, it makes sense to say that while the assertion 'That is a dog' is purely descriptive, the assertion 'That is a good dog' is half-evaluative, half-descriptive, whereas the statement 'Dogs are good' is

²⁷ In contrast, I read another prominent commentator Jeremy Wanderer (2008) to urge Brandom to instead hold onto normative phenomenalism over phenomenalism about norms, thereby preserving his realist stance on norms and meaning, although at the cost of downgrading the ambitions of the theory from explanatory to elucidatory ones. These competing readings will be elaborated further in Sections 10.1 and 10.2.

purely evaluative (albeit perhaps vaguely so). Most importantly, the same term token cannot be used both descriptively and evaluatively at the same time: it must be the case that the use reports an objective fact or a subjective attitude, but not both simultaneously.

The first thing to notice is that there is an important exception to the rule that obliges some separate discussion. Namely, pejorative terms, such as 'mudblood' purport to report a fact (the target is an offspring of a mixed union between a muggle and a wizard/witch in J.K. Rowling's *Harry Potter* fictional novels) *and* express a derogatory attitude towards them.

One popular response by philosophers has been to argue on the intuition that slurs and other derogatory terms are not only obviously morally but also semantically suspect, precisely because they mix the descriptive and evaluative functions in a single term. One promising proposal along these lines is made by Dummett (1993), who argues that slurs violate the *harmony* condition on introducing new terms into a language. Briefly, introducing 'mudblood' into a language entails creating new inferences between claims that were not previously possible: the term represents a non-logical case of a non-conservative extension to language.²⁸

It is now interesting to observe that Brandom objects to Dummett's formal criticism of slurs based on his material inferential understanding of content, according to which slurs are not *semantically* any more suspect than terms, such as 'cruel', are in two ways. The first reason has to do with his Hegelian idea that determinate conceptual contents must be understood via the process of determining them via applications. To demand that non-logical words of a language must have conservative extensions, in that adding new words or claims does not change the content of old ones, represents a 'peculiar ideal' where all the existing contents would be 'out in the open' and transparent to their possessors, who could never be truly surprised by their own meanings (MIE, 127). In hindsight, it could be that already, Brandom objects to Dummett based on his broader commitments regarding the

²⁸ The original idea of harmony was introduced by Nuel Belnap (1962) in response to a problem in logic presented by Arthur Prior (1960).

nature of conceptual contents, criticising him for subscribing to the *Verstand* metaview of concepts instead of *Vernunft*.²⁹

The other reason to criticise a formalist rejection of slurs is that these terms are not faulty by joining otherwise heterogenous functions (descriptive and evaluative) in one term. In comparison, to call something 'cruel' is to join descriptive circumstances of application (e.g. 'intentionally harming someone unable to defend themselves') with evaluative consequences of application ('ought not to be done' or 'is bad'). The ultimate reason for rejecting slurs like 'mudblood' is not formal but material, namely that one rejects the material inference itself as invalid, because it is not true that being born of a mixed union between a wizard/witch and a muggle would merit condescending or hostile attitudes.

Returning to our original concern of how it is possible that meanings of words could change retrospectively in time, we can first see that while some terms are special³⁰ in mixing descriptive functions with evaluative ones at the level of *semantics*, *all* terms share the trait of having evaluative, normative consequences at the level of their *pragmatics*. The mutual temporal recognition model suggests that when a judge applies a concept to a case, they assert something and expect future judges to acknowledge and accept it. Importantly, not only is petitioned recognition for the applicability of the implicit norm chosen by the judge, but also the authority of the judge herself in applying the norm. The difference between the authority of the norm itself merely conveyed by the judge's own authority just is the difference between the semantic and the pragmatic dimensions, or in more familiar terminology, the difference between the word meaning and the speaker meaning. Crucially, *if* Brandom can show that the authority (applicability) of the norm (meaning) itself can be changed by changing the authority accorded to the present judge herself by the future ones, *then* it will follow that norms and meaning can change retroactively in time.

²⁹ Brandom does agree with the idea that specifically logical words should have conservative, harmonious extensions.

³⁰ In fact, it may well be that those words, which do *not* mix the descriptive conditions of application with the evaluative consequences of application, are in the minority. For even a seemingly value-neutral word like 'dog' could be seen to have as its content the material-inferential consequence 'makes for a good pet'.

That is as far as I am able at this point to make intelligible the quirky semantic temporality resulting from Brandom's bold synthesis in its advanced form. I have not yet shown, however, how Brandom is committed to the quirky temporality, or how he himself develops the Hegelian mutual recognition account in *A Spirit of Trust*. That shall have to wait until Chapter 3.

9.4 Expressive Algorithms

At this point, it is good to have a brief reminder about semantic dispositionalism, as the doctrine was discussed in Chapter 1. One important dispositionalist strategy was called the 'algorithmic response', according to which the meanings of words are determined by a kind of algorithm possessed by the subjects as a (mental or physical) dispositional state. Now, it might be surprising to notice that Brandom, who otherwise is just about as far from semantic dispositionalism on the theoretical spectrum as anyone can be, has reserved an important place for algorithmic abilities in his post-MIE theory. In this subsection, my aim is to explain how Brandom's use of algorithmic abilities differs from semantic dispositionalism, and by that token, shed light on his brand of semantic expressivism.

The major difference can be thus summarised: where dispositionalism attempts to use algorithmic abilities to *describe* meaning-determining facts, for Brandom, algorithmic abilities serve a special *expressive* role. In this sense, the use of algorithmic abilities is fully in line with his semantic expressivism, which seeks to explain the significance of semantic vocabulary not in terms of its descriptive function but its expressive function. In his words, semantic vocabulary, such as the locution, '...means that...' enables us to *say* something, which we could previously only *do*, not to describe some special realm of semantic facts.

In Section 9.2, we already saw *what* semantic vocabulary expresses, namely the implicit material inferential significance of assertions within the scorekeeping practice. Here, I shall focus on describing the expressive relation itself, or how the process of 'making explicit' should be understood.

To begin with, the specifically semantic 'making explicit' cannot be understood as a purely *formal* transition from an 'implicit state' where the propositional content expressed would itself remain the same: semantic 'making explicit' is not merely *explication*. For reasons that are best left to be discussed later in Section 10.1, 'making explicit', which is a kind of a doing, must be understood as somehow 'tampering' with the sayable content itself.

It is crucial now to emphasise that this remark about specifically 'semantic making explicit' does not necessarily extend to other ways one can make vocabularies explicit. For one, according to Brandom, logical vocabulary must *not* be understood as 'tampering' with the contents it makes explicit in specifically logical form. Underlining this observation is the point that Brandom is a *local* expressivist; a position that can be summarised as follows:

- a) Natural discourse is formed of distinct *vocabularies*, such as semantic, indexical, modal, logical, etc.
- b) The vocabularies are made distinct in part by the different *functional tasks* they have (been) developed to perform.
- c) Although the functional tasks performed can all be broadly characterised as *expressive*, for each distinct vocabulary corresponds a distinct *way* in which the generic expressive function operates. There is no *universal* expressive 'making explicit' relation across all possible vocabularies.

In contrast, *global* expressivism, as pursued perhaps most notably by Huw Price (2013), denies that (c) would hold, although it is otherwise amenable to the idea that natural discourse can be demarcated in different sects of vocabularies in some broadly functionalist sense. The comparison between the respective merits and problems of local and global expressivism would be a book-length topic and some remarkable efforts on this front have already been done, e.g. by Pietro Salis (2019). Here, I shall not undertake an examination of the comparison, but focus solely on expressivism as Brandom advances the tradition, especially regarding the semantic and normative vocabularies. To do that, we must look at the ideas (a)-(c) a little more closely.

First, what is a 'vocabulary'? As far as I know, there is no definition offered for the term anywhere in Brandom's extensive oeuvre, although I cannot say the same about Rorty, from whom Brandom inherits it (PP, 117). Perhaps the closest thing to a definition would be 'something that is *said*', which goes to show that 'vocabulary' is best treated as one of those words philosophers use to get started but rarely look back on once they get going.

We get a better idea about what vocabularies are by seeing how and what they are used for. Brandom is very fond of later Wittgenstein's many metaphors about language, especially the one where he makes an association with a toolbox (PI, 11§). An important way in which various tools can be classified is based on their functional tasks, e.g. attaching things together (nails, glue), separating things (saw, crowbar), and so on. Similarly, the traditional thought in analytic philosophy has been that all genuine vocabularies³¹ can be classified based on one key task function, which is that of *describing* things. So, modal, normative, semantic, and logical vocabularies (i.e. what is *said* to be modal, normative, semantic, or logical) are understood as describing the respective classes of *facts*, as well as coding the respective categories of *knowledge*. The contrasting idea that Brandom picks up from later Wittgenstein and Sellars is that for all these and more vocabularies, *description* is not the task function that can be used to explain the proprieties of their use, for these are *expressive* vocabularies. And even as broadly expressive vocabularies, they must be thought of as performing the function in varied, not clearly comparable ways, just like nails and glue must be, although both serve the generic functional task of attaching things together (BSD, 3-7).

The problem with Wittgenstein's metaphors and analogies has always been to make them precise enough to be useful beyond vague illustration. Conveniently, this is exactly what Brandom's admirable technical rigour is poised to do: the resulting

³¹ By 'genuine' vocabulary, I mean a vocabulary the sentences of which really have meaning according to the theory in question. For example, according to the traditional analytic philosophy of Rudolf Carnap (1931), metaphysical (especially the kind found in *Sein und Zeit*) vocabulary is ingenuine because it is meaningless, and meaningless because it cannot be descriptively grounded in empirical-phenomenalist vocabulary, which is taken as the foundation of all meaningful discourse. That is not to say that metaphysics is active nonsense, according to Carnap, only that proprieties governing its use cannot be understood as descriptive but rather as expressive (of emotions, say) and thus akin to (bad) poetry.

research program he aptly calls 'analytic pragmatism'. The central questions of (and about) this program that I shall focus on examining here are:

- 1.) If traditional analytic philosophy understands the proprieties of use to consist of descriptive proprieties in case of all genuine vocabularies, how is the notion of *expressive proprieties* to be described?
- 2.) How are the expressive proprieties regarding what is *said* to be understood based on the practical abilities to *do* something?
- 3.) What are the main *theoretical advantages* of expressive proprieties of use in relation to the descriptive proprieties of use in case of, e.g. modal, logical and normative vocabularies?

Answering the first question will be the easiest since we have already done much of the work needed to understand it. Briefly, whereas the descriptive proprieties are readily understood in terms of such extensional concepts as truth, representation, or reference, the expressive proprieties that Brandom aims to supplant them with primarily consist of material inferential relations.

What is the novel element that distinguishes analytic pragmatism as a distinct research program from being a new name for semantic pragmatism? Although the doctrines are *de facto* interrelated simply because Brandom is their common paragon, as a more general research program, analytic pragmatism is not committed to the existence of meanings, unlike semantic pragmatism. Because semantic pragmatism stands for the foundationalist claim that what joins expressions with their meanings is the use made of those expressions, the theory presumes that there *are* meanings to be joined with expressions to begin with. In contrast, analytic pragmatism is more a *method* than a theory or claim about meanings, and in principle, it is compatible with semantic eliminativism. Thus, the 'meaning-use analysis' to be elaborated below can discuss 'pragmatically mediated semantic relations' between different vocabularies without the implied commitment that any of these vocabularies are genuinely meaningful in the sense of footnote 31 above, or of what exactly the meaningfulness consists in (to wit, whether the pragmatic relations need to be given a normative spin).

We can now move on to discuss question number (2), which takes us to examining Brandom's meaning-use analysis. To start off, the meaning-use analysis operates on three basic kinds of meaning-use relations (MURs):

PV-sufficiency: A set of practices or abilities P is sufficient to *deploy* vocabulary V .

VP-sufficiency: A vocabulary V is sufficient to *specify* a set of practices or abilities P .

PP-sufficiency: A set of practices or abilities P is sufficient to deploy *in principle* a set of practices or abilities P' .

It is good to focus first on PP-sufficiency relation, for that is where algorithmic abilities come into play. The 'in principle' way in which one set of practices or abilities suffices to deploy another is understood by Brandom in the algorithmic sense in which, e.g. the ability to do long division just consists in the ability to do multiplication and subtraction according to a certain branched-schedule algorithm (BSD, 26). The founding idea of the meaning-use analysis is that our ability to deploy various vocabularies, and by that token, the phenomenon of natural discourse itself, can be understood on the model of such *algorithmic elaboration*. Once more, the key cue is picked up from Wittgenstein, who asks us to imagine a community who knows how to give names to people, and then learns how to give names to rivers. What makes this generic account of conceptual projection 'pragmatic' is that it is, at the outset, considered being *only* a question of *contingent social, historical, and evolutionary fact* of how the *practice* of giving names to people can be extended to cover, e.g. rivers (BSD, 6).

I am not going to expound on how the technical detail of the meaning-use analysis is supposed to work, i.e. how to understand the 'pragmatically mediated semantic relation' between two vocabularies (the VV-sufficiency MUR) as composed of a VP-sufficiency, PP-sufficiency and PV-sufficiency claims. Neither will I touch upon how Brandom argues for his local expressivism by showing that although different VV-sufficiency relations will be underlined by the meaning-use diagrams of a similar *structure*, the expressive relation itself need not be given a uniform interpretation

across the board. Instead, after briefly finishing this section by studying what theoretical advantages analytic pragmatism has over the classical project of analysis, I shall in the next section focus on perhaps the most important expressive, 'making explicit' relation in Brandom's works, namely the expression of norms implicit in the practices by the 'means that' locution.

There are two basic advantages that analytic pragmatism and its companion commitments, such as semantic and methodological pragmatism, enjoy over: on the one hand, classical analysis and, on the other, the more recent externalist semantic projects, such as the causal-historical theory of reference. Both advantages relate respectively to the order and means of explanation. To begin with, the traditional project of analysis takes the descriptive task-function to be the prime directive, according to which the genuineness of all vocabularies is to be measured. So, it is presumed at the outset that if a vocabulary is genuinely meaningful, it must correspond (or represent) some distinct class of facts (or a distinct category of entities), of which the subjects can have distinct kinds of propositional knowledge. In other words, the semantic strategy opted for through the classical analysis ties it at the outset either to substantial metaphysics and epistemology in case of, e.g. modal and normative vocabulary, or then to severely limiting the space of genuinely meaningful vocabularies, e.g. perhaps only to those which can be logically derived from the descriptive empirical-phenomenalist vocabulary.³²

It is interesting to observe that the order of explanation where metaphysics (or ontology), together with epistemology, are supposed to do crucial explanatory semantic and ultimately pragmatic work ('pragmatic' in the sense of determining the proprieties of use) is broadly the same as what Kripkenstein attributes to Low-brow Meaning Determinism. For MD_{LB} precisely attempts to explain the proprieties of Jones's behaviour with the '+' sign by his knowledge and the semantic facts of the matter. This is the general sign of what is called an 'intellectualist' order of explanation, which is opposed to the pragmatic order of explanation favoured by Brandom.

³² The 'empirical-phenomenalist' descriptive vocabulary is not a monolith, but comes with different degrees, which, e.g. Sellars, analyses in his work (1956).

But what is the *advantage* of reversing the order of explanation? The main reason is that the practical order of explanation is more 'down to earth', both on the side of the world (metaphysics and ontology) and on the side of the subjects' epistemology. Already in MIE, Brandom congratulates Russell, whose classic thought led him to postulate negative facts and conditional facts to explain the contentfulness of logical vocabulary, for his 'conceptual heroism', if not for his 'good sense' (MIE, 76). Following this route leads to proliferating one's ontology for ultimately non-ontological reasons, which arguably puts the cart before the horse. The corresponding idea rejected on the side of the subjects' epistemology is that (literally) too much is required of speaking beings to have propositional knowledge of the facts to which their fluency in a vocabulary is to be due. The general argument for why 'knowing-that' cannot be postulated to explain every practical piece of knowing-how to apply a vocabulary is pursued under the criticism of regulism (MIE, 20); it is of a piece with Wittgenstein's regress of interpretation and Lewis Carroll's 'Achilles and Tortoise' arguments.

The reason analytic pragmatism is more down to earth than the classical project is that it can do semantic work without committing at the outset to substantial metaphysics and the proliferation of different kinds of facts regarding realms, such as modality and logic. In fact, all that Brandom claims is required to get the algorithmic elaboration started is an 'autonomous discursive practice' (ADP), which means a practice in which the subjects can make implicit assertions and material inferences; in other words, the scorekeeping practice (BSD, 42). Once subjects can engage in the scorekeeping practice, they already implicitly know-*how* to deploy, e.g. the vocabulary of indexical expressions ('I', 'you', etc.), and that by suitable algorithmic elaboration, they can learn to express their implicit know-how in explicit sayings-that (BSD, 57). Brandom calls such a procedure 'pragmatic bootstrapping'.

The phenomenon of pragmatic bootstrapping enables Brandom to make flesh another one of Wittgenstein's famous metaphors, namely of language as a city:

Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses. (PI, 18§)

The centre of the city, for Brandom, is the practice of making assertions and inferences, around which the suburbs of various 'add-on' vocabularies arise in the flux and flow of contingent socio-historical evolution. The secondary literature has also grown accustomed to calling this the 'layer-cake' picture of language and rationality, although I cannot say why.³³

In sum, the advantage of analytic pragmatism over classical analysis is to be more down to earth, while also keeping firm technical rigour. To be more down to earth means, first, to be less metaphysically, ontologically, and epistemically committed than classic analysis and, second, being in a more opportune position to do homage to the broader context in which language is seen as a socio-historical, evolutionary phenomenon that did not spring up from the ground all at once but has contingently developed the way it has to this day.

Of course, the classical project of the analysis has not been the only contender in analytic philosophy for a while now, nor even the most popular. How does analytic pragmatism fare against the formidable Kripkean referential tradition, for instance?

To start off, as discussed in Chapter I, I believe that the (by now) classic Kripkean theory of reference offers a promising candidate to advance the lessons taught by Kripke's sceptical challenge about how meanings are determined. Moreover, I think the account is comparable with Brandom's in several ways; one important point concerns modality. For one central reason by which Kripke can criticise the classical project of analysis focused on the descriptive task function in *Naming and Necessity* is by making free use of alethic modal notions, most importantly, rigid designators and counterfactual possible worlds. By 'free use', I mean that aside from the fearsome expressive power due to the technical rigour of his modal logic, Kripke does not so much as worry about what (primarily empirical) *justifications* there can be for deploying the modal vocabulary, hence endorsing modal truths and knowledge. The absence of justification for deploying modal vocabulary is precisely what

³³ One reason might be that it is more comfortable to cut pieces of a cake rather than bulldoze down suburbs. For one strand of critics sympathetic to Brandom's pragmatist expressivism object to the particular structure of the cake and what is indeed essential to its base/centre and not just a contingent historical add-on. Especially the contingency of logical vocabulary (Lauer 2012) and the indexical vocabulary (Lance and Kukla 2009; Wanderer 2021) have been discussed in the literature.

Brandom criticises Kripke of, in the immortal, paraphrased words of Alexander Pope:

Modality and Nature's laws lay hid in night, God said: "Let Kripke be!" and all was light. (BSD, 94)

In contrast, we have already seen how Brandom appeals rather to *deontic* modal vocabulary, most importantly, that of discursive commitments and material incompatibility relations, in his technical apparatus. However, not only does Brandom have a philosophical justification for his use of deontic technical primitives in the form of his scorekeeping model, but he also claims that he can justify the use of alethic modal vocabulary in terms of the model under the heading of the 'Kant-Sellars' modal thesis (BSD, 96). It should come as no surprise that the Kant-Sellars modal thesis can be formulated by using the meaning-use analysis: the gist of the thesis is that if one can participate in an ADP that includes ordinary empirical descriptive terms, such as 'red' and 'mass', one already knows-how to do everything needed to deploy the alethic modal vocabulary of (empirical) necessity and possibility.

9.5 What Are Norms Implicit in Practices?

We have now seen the basic architecture of Brandom's answer to the question of what meaning is: it is a certain socio-historico-perspectival process of administering material inferential significance known as scorekeeping that consists of expressive algorithmic abilities to elaborate and explicate the said material proprieties. Meaning is the product of our rational, inferential, linguistic, expressive algorithmic practices, and abilities; a claim that aims to balance between the realist and anti-realist understandings of 'product'. According to Dummettian, Davidsonian, or Quinean anti-realism (or eliminativism), meaning is nothing over and above what subjects in the practice can come to *understand* as meaningful in their communal interactions.³⁴

³⁴ These authors differ, of course, massively in a variety of other related issues, and there is some internal vacillation among at least Davidson and Dummett through their careers. But in regard to meanings, or propositional content in the intensional sense, all three would agree in their mature views that it makes no sense to distinguish between facts about meaning and facts about what someone, or everyone, can understand as meaningful. Whatever differences among subjects arise in terms of

This is the insight Brandom seeks to consider under the programmatic heading of phenomenalism about norms. On the realist side and under the demands of normative phenomenalism, Brandom maintains that while meaning is the product of our discursive abilities and practices, it is not reducible to the practices as *immediately* understood by the subjects within it. Meaning has a kind of genuine autonomy or authority over the attitudes in that facts about meaning do not supervene on facts about attitudes; meaning is *real* in a stronger sense than what the anti-realists presume.

The aim of this section is, first, to illustrate Brandom's *motivations* for holding onto the ambitious task of synthesising the realist and anti-realist facets of his theory. Second, I shall further try to specify what exactly the *criteria of success* for the synthesis are, or how phenomenalism about norms clashes with normative phenomenalism. The order of these tasks is not accidental, for I think that without strong motivations driving it, it is not altogether clear why the synthesis would be necessary or even desirable over and above choosing between the individual appeal of the respective original commitments. The reason to pursue the synthesis must be something else than the combined appeal of the original commitments, for all things being equal, it seems more prudent to accept the simplest explanation and favour the anti-realist appeal over the realist alternative and vice versa, but not both at once, no matter how theoretically satisfying that might be.

In my view, the most powerful motivation that Brandom offers for pursuing the synthesis is that it is a plausible way to give an answer to three interlocking problems relating to natural language: the *emergence problem*, the *demarcation problem*, and the *leverage problem*. The emergence problem concerns the question of how natural language developed in the course of natural history. The demarcation problem concerns the question of what distinguishes non-linguistic and non-discursive animals from linguistic and discursive humans. The leverage problem concerns the question of how to explain the evident, tangible difference in the power level of expressive abilities between beasts and men, or 'the bonanza of new abilities and possibilities that language opens up' (PP, 28).

communication, for instance, must be explicable either in extensionalist terms or then by misunderstanding, but not by 'meanings themselves'.

There is a dynamism between the emergence and leverage problems, such that any independent answer to either makes responding to the latter that much more difficult. For example, Brandom criticises Dewey for committing the mistake opposite to the 'Platonist Cartesian', who has a ready explanation for the leverage problem (namely that only we *can* think thoughts), of not adequately emphasising the drastic difference in abilities that having language implies, e.g. the sheer number and complexity of thoughts a speaking being can entertain relative to a non-speaking one. While it is indubitable that linguistic creatures have to be placed on an evolutionary continuum from non-linguistic ones, it is equally necessary that the actual expressive abilities that language implies are captured in their true 'bonanza' as Brandom puts it.³⁵ The true problem is to offer a combined solution to the emergence and leverage problems, which is precisely what the demarcation question is poised to do. For the demarcational question is meant to deliver the answer to what is missing in non-linguistic animals that humans enjoy in their bonanza (PP, 29).

Why is it, then, that a synthesis between semantic realism and anti-realism is required for answering the demarcation question? As far as I can discern, the two main reasons that Brandom accepts are, first, the fundamental pragmatist idea that whatever meaning is, it must be an active product of our practices and abilities – not just a passive reflection of the autonomously existing world or conceptual realm explicable by representationalist resources. Second, he seems to think that while the pragmatist-expressivist-phenomenalist account offers a useful starting point to metasemantics, we must not thereby think that propositional contents, what is claimed and asserted, can be subjugated to the acts of claiming and asserting without a remainder. In the end, Brandom takes it that the Platonic Cartesian has a better answer to the leverage problem than, e.g. Dewey, ever did, namely that humans have special expressive powers because *we can think genuinely intentional thoughts* while the animal cannot. What the Platonist lacks is a proper understanding of what the power

³⁵ It is far from trivial how to actually describe the bonanza. Here, I shall only limit to two qualifications that I think are most crucial for Brandom. On the qualitative side, the expressive power of language is special in virtue of inaugurating access to specifically propositional contents, i.e. contents one can *reason* with. Indeed, Brandom often speaks of ADP as if it represented rationality as a natural (socio-historico-natural?) kind. On the quantitative side, Brandom puts a lot of emphasis on the compositionality (both syntactic and semantic) of language, which means the exponential growth of possible novel sentences (by that, token thoughts) that a speaking being has access to in virtue of language.

of thinking consists in, namely its roots in language, and most importantly, what constraints on the contents of thought are thereby implied. Nonetheless, to give up a robust notion of propositional content would make the leverage problem too difficult to answer.

So, here we have the motivation for pursuing the grand synthesis. How credible is it? Considering how rare are the attempts to try to tell a coherent, detailed story about the emergence, demarcation, and leverage problems at once, I believe the prize of the added complexity is certainly worth it.

The question then becomes: how is Brandom able to pull off the grand synthesis, thus solving the demarcation problem, as well as the emergence and leverage problems? The key to all this is the idea that genuine discursive norms may be 'implicitly instituted' by a certain socio-historical practice. The norms 'confer' objective, propositional contents on assertions and material inferences of the practitioners. To end this section, I shall broadly characterise the core problem with the idea of implicit institution of norms before moving on to examine the more detailed criticism Brandom has faced over the years.

The theoretical space for implicit norms in *Making It Explicit* is first carved between two critical arguments about rules, respectively known as refutations of *regulism* and *regularism* (MIE, Ch.1/III). Briefly described, regulism claims that all norms are to be understood as explicit principles, while regularism claims that rule-following can be understood to consist of the regularities of behaviour.³⁶ The combined refutation of these claims implies that if there are any *genuine* norms at all, some of them must be implicit in the sense that they neither presuppose explicit articulation in principles, nor are reducible to regular behavioural patterns. I shall now examine this chain of arguments in more detail.

Brandom defines regulism as follows: "This view, that proprieties of *practice* are always and everywhere to be conceived as expressions of the bindingness of underlying principles, may be called *regulism* about norms' (MIE, 20). Brandom names Kant and

³⁶ In Brandom's terminology, 'rule' is an explicit form of a norm, so the denial of regulism amounts to denying that all norms would be rules. In this work, I have opted to use 'rule' and 'norm' interchangeably, marking the distinction between explicit and implicit norms otherwise.

Frege as historical figures of this 'Platonic' account of norms, which takes normativity as it appears in written law as a general model.

The knock-down argument against regulism Brandom attributes to Wittgenstein's *Investigations*, although he notes that Lewis Carroll made a closely parallel point about logic already in his work (1895). The point is that every rule has normative significance, i.e. it may differentiate between what is correct and incorrect according to it, only when it is applied in practice; yet, its application itself cannot be understood as governed by principles because this would lead to a regression. Carroll argued that not all rules of logical inference can be expressed in axiomatic form, for if the application of the axioms itself demands preexisting axioms, the regress is inevitable. According to Brandom, Wittgenstein makes the further claim that all explicit norms owe their normative significance, i.e. their ability to distinguish what is correct and incorrect according to the norm, to practical, implicit normativity (MIE, 22).

Onto regularism then. The problem left in the wake of rejecting regularism is evident: how is the intuitively plausible notion of rule-governed practice possible if norms cannot be understood all the way down as explicitly codified principles? What is the nature of the relation between the rule and the practice, and how is it possible to apply a rule meaningfully, i.e. in Brandom's terms *with justification*? This is the fundamental normative condition on meaningful use, which Brandom thinks must be preserved by an adequate pragmatist account of language, his main reason for rejecting both regulism and regularism.

To begin with, 'simple regularity theory' claims that implicit rule-following just is regular behaviour, i.e. behaviour governed by some causal or at least statistical regularity. The subject herself need not be able to understand oneself as following any rules for the theorist to understand her as doing so by conforming to the regular patterns of behaviour. The major argument against regularism that Brandom attributes to later Wittgenstein is that it makes the difference between correct and incorrect application of a rule unintelligible. By losing this distinction, we also lose

another seminal one for understanding the meaningful use of language: what *is* actually done and what *ought* to be done³⁷ (MIE, 27).

The argument why at least simple regularism cannot make this difference is close to the criticism that Kripkenstein launches, and Brandom agrees that Kripke is on point here. Regularism is committed to understanding the difference between the correct and incorrect use of signs by the difference between regular and irregular behaviour. Yet, any set of actual behaviour exhibits indefinite patterns of regularity and irregularity both, for regularity is a relative notion. There must be some standard, according to which regularity, e.g. of a clock, can be measured for claims about its regularity to have (substantial) truth conditions. The refined proposal by regularism is that it is the semantic dispositions of the subject that set such a standard; an argument the merits of which were discussed in the previous chapter, and which Brandom settles describing as 'controversial'. Here, it suffices to focus on the problem that Brandom sees as even more pressing for regularism/dispositionalism:

Understanding the norms implicit in practice as descriptively adequate rules codifying regularities of dispositions (even if a unique set of such rules is forthcoming) loses the contrast between correct and mistaken performance that is of the essence of the sort of normative assessment being reconstructed. If whatever one is disposed to do counts for that reason as right, then the distinction of right and wrong, and so all normative force, has been lost. (MIE, 29)

The basic problem with this setup that introduces us to implicitly instituted norms is that nowhere do we get a positive, direct argument for their existence. What we have instead are two negative arguments, convincing by themselves, which, together, set a lower and a higher limit to the possibility of genuine normativity. Regularism is the higher limit since its criticism shows that not all norms can be understood by explicit rules; regularism is the lower limit, since no genuine norm can be understood to end in dispositions. So, how do we get the *positive* claim that there *is* something between these limits to which the concept 'implicit normativity' applies? The kicker is that *unless* there is something there, genuine normativity becomes conceptually impossible under the conditions we have already accepted. And since Brandom accepts at the outset that there *are* genuinely normative phenomena, most

³⁷ Later on, I shall refer to this idea as the 'is-seems' distinction: what is correct cannot be identified with what is taken (or what seems) as correct.

importantly, the very faculty of our reason, the conclusion follows that there must also be implicit norms.

Of course, there is no place for one philosopher's *modus tollens* than another's *modus ponens*. Consequently, the critics who have found Brandom's synthesis wanting have either concluded that the scorekeeping model is not really an *explanation* of language (in the sense of resolving the demarcation problem), or then that it *explains away* troublesome notions, such as meaning and genuine normativity. The following sections (10-10.3) elaborate on this dynamic further by mechanically going through what I consider being the most important critical arguments against Brandom's grand synthesis between normative phenomenalism and phenomenalism about norms. These will then serve as the contrast against which I make my own contributions, first in Sections 11 and 12 and then in Chapter III.

10 CRITICISM OF THE SYNTHESIS

One way to criticise a synthesis is to demonstrate that it leans towards one thesis rather than finding a balanced approach. Of course, to show this is not to show that Brandom's project in MIE fails as such, only that it fails as a synthesis. This attitude has been, I believe, dominant in the reception of MIE and other Brandom's works tightly connected to it. The critics can thus be classified, according to which part of the synthesis they take should prevail over the other. I shall call these 'the normative phenomenism camp' and 'the phenomenism about norms camp' respectively. These groupings are loosely formed because normative phenomenism and phenomenism about norms are programmatic commitments that are sensitive to context. Nonetheless, I think there is a clear trend to how the critics approach Brandom from either side of these two limiting ideas. This classification cannot cover all critics because not everyone agrees with the two programmatic commitments.

I start by discussing the criticism of the normative phenomenist camp, who take Brandom's semantic realist side to be the most important one. A characteristic of this camp is wariness of Brandom's idea that discursive normativity could be *explained* in terms of non-normative explananda, which would come very close to the idea of reducing it to something non-normative metaphysically speaking. As a result, the camp is critical especially of Brandom's layer-cake picture of rationality, claiming that the base is ultimately not strong enough to sustain genuine normativity or objective enough conceptual content.

10.1 The Cake Is a Lie

One of the earliest formulations of the core synthesising problem in MIE comes from Gideon Rosen (1997) in two steps. First, he observes Brandom's commitment to the claim that discursive, conceptual norms are explainable based on more primitive kinds of practical, non-conceptual norms that are only implicit in the

practices of a pre-discursive community, i.e. the community that causally and conceptually precedes the merely rational (i.e. not yet logical) centre of ADP. In what follows, I shall call this more primitive community the 'proto-hominid community'.

Second, Rosen notes that, according to Brandom, we can adequately describe the pre-discursive practices where no norms are made explicit in non-normative terms, e.g. by the reliable differential responsive disposition (RDRD) vocabulary of issuing external sanctions. The problem is that if the pre-discursive practices of the proto-hominids suffice to explain the emergence of discursive, conceptual ones, and if the pre-discursive practice is adequately described in non-normative dispositionalist terms, how is it that the more developed conceptual practice would not be? Yet, that kind of reduction is precisely what Brandom's commitment to normative phenomenalism denies. It follows that either the conceptual practice of ADP is not genuinely normative, or else that it cannot arise independently from the intentional descriptions of the proto-hominids' practice. The latter conclusion is what Rosen surmises is true, hence that, *contra* Brandom, the intentional cannot be explained *by* the normative but *with* it. The moral that Rosen drew, later echoed by other critics, is that if the intentional cannot be reductively explained by the normative as Brandom's main claim goes, the remaining option where intentional and normative vocabularies are treated as equal primitives can still count as philosophically 'explicative'. Although I will not stop here to examine what 'explication' means exactly, the point will recur later in the story.

An iteration of Rosen's critical thought comes from Anandi Hattiangadi (2003). The (abbreviated) explanatory order she attributes to Brandom is that normative statuses are explained by normative attitudes, and normative attitudes are explained by non-normative dispositions to sanction behaviour, which poses the question of how genuine norms can then arise. In particular, how is it that the practices that are adequately described in merely dispositional terms suffice to determine the contents of the discursive, conceptual norms, given that any set of dispositions conforms with an indefinite number of norms. The only solution (apart from un-Brandomian quietism) that Hattiangadi sees available here is a reversal of the idea that normative attitudes institute normative statuses:

The idea seems to be that the normative status of the actions of the [pre-discursive community] is a function of *our* attribution of that status: *we*, so to speak, make it

possible for them to institute their norms by attributing to them the ability to attribute correctly or incorrectly. So one more level of explanation needs to be added to Brandom's picture. As it is, we had normative status explained in terms of attributions of normative status, attributions explained in terms of practical deontic attitudes, and these attitudes explained in terms of dispositions to sanction. Now, we need to explain the practical deontic attitudes in terms, not just of the natural abilities of the members of the community, but *also*, in terms of normative statuses *attributed* explicitly by us, the interpreters in this exercise. (Hattiangadi 2003, 428)

If our attribution of status is necessary to discriminate *which* commitments and entitlements are being undertaken or attributed by the fictional [pre-discursive community], then it will turn out that *explicit* attribution of normative status (by us) is necessary for the *implicit* practice to be one capable of mere normative sanction. But this has Brandom's order of explanation back to front. Of course, we need to use our explicit vocabulary to talk about the [pre-discursive community], but that talk cannot be a *necessary* condition for practices to implicitly institute norms. (Hattiangadi 2003, 429)

The essential critical point that both Rosen and Hattiangadi make is that the only way in which the pre-discursive practices may 'institute' discursive ones is that we allow either some explicit normative descriptions or intentional ones of the pre-discursive practice to determine *which* commitments they have *really* committed themselves to. However, in that case, recourse is made to explicitly normative or intentional vocabulary, which leads either to the circularity of the account (Hattiangadi) or then significantly weakens its explanatory scope to a 'merely' explicatory one (Rosen). In either case, the pragmatic bootstrapping feat is not explanatory in the strong sense, and the 'collapse of external perspective to internal' fails. For the 'collapse' to succeed, the community that institutes discursive norms must exhibit 'original intentionality', i.e. intentionality that an external interpreter can correctly attribute as instituted only by the community's own attitudes. More or less, the same circularity argument as what Hattiangadi levies against Brandom is made, e.g. by Glüer and Wikforss (2009, 62-63); see also, Hattiangadi (2006, 235).

Daniel Laurier (2005) points to two further, similar yet distinct circularity problems in MIE. While Rosen and Hattiangadi trace the circularity from the implicit to the explicit, arguing that the determination of norms must involve explicitly normative vocabulary, Laurier, more straightforwardly, observes that:

an account such as Brandom's can be expressively complete only if it is circular! Expressive completeness requires that it be possible to *say*, i.e., to express explicitly, that someone practically attributes a given discursive deontic status, for example, that

someone attributes a doxastic commitment to it's [sic] being the case that *p*. But if what is thus said is conceptually contentful (as it should be) and truly reports what one is doing when one practically attributes a given doxastic commitment, then practical attributions must themselves be conceptually contentful. For what is practically attributed must then be the same as what is (truly) *said* to be practically attributed. It follows that if the proposed account of content is expressively complete, it is circular—since the practical deontic attitudes must then be conceptually contentful. (Laurier 2005, 149)

The problem with Brandom's metaphors of 'conferral' and 'institution', descriptive of the process of making explicit here is that if 'making explicit' is understood as *explication* that must be either true or false (correct or incorrect), then surely, if it is to be true, what is being explicated must *already* be conceptually contentful. But in that case, it becomes difficult to see how anything is 'instituted' in between.

The other circularity problem that Laurier identifies follows from an attempt to charitably remedy the obviously devastating first problem. The process of 'making explicit' cannot be interpreted as formal explication, which would add nothing substantial to what is made explicit (i.e. the shift from *doing* to *saying* cannot here be purely a matter of form), for the resulting circularity would be too obvious. Moreover, several paragraphs in MIE speak against that interpretation. However, it is not clear how the relations between the implicit and explicit should be understood considering Brandom's contention that there could be 'merely rational' subjects who could implicitly attribute not only normative statuses but also normative attitudes to each other, although they could not make the attributions of either explicit. The reason this is problematic is that Brandom clearly says (MIE, 499) that the ascriptional, explicit locutions are necessary for distinguishing between the status of being committed to *p* and the attitude of acknowledging/attribution a commitment to *p*. Since the 'merely rational' practice cannot include explicit ascriptional locutions, the subjects therein assumedly cannot differentiate between attributing a status and attributing an attitude. But in that case, there is no way for the merely rational, implicit practice to involve any conceptual content, which is essentially perspectival for Brandom, by involving the distinction between what is taken to be correct and what is correct, i.e. the attribution of status and attitude, respectively. It follows that:

[i]f this is right, then it looks as if Brandom would have to give up either the claim that ascriptional locutions are optional and that there may be 'merely rational' agents, or his account of what the institution of deontic statuses and objective conceptual contents consists in (or at least that part of his account which calls for rejection of

the idea that merely rational agents are capable of higher-order practical attitudes). (Laurier 2005, 152)

Laurier assumes that the first option is correct. Next, he looks at the fallibilist, structural social-perspectival solution, according to which the distinction between what is correct and what is taken to be correct can and must be drawn from every deontic perspective, so that although there is never a final answer to what is 'really' correct, the structural notion of objective normative statuses persists. Laurier points out that:

[e]ven if it can be granted that everyone involved in a discursive practice must be able to contrast what is correct with what is taken as correct by someone else, and what someone else is 'really' committed to with what he merely takes himself to be committed to, this can at most sustain the conclusion that every rational agent has some (perhaps only practical) grasp of some notion of objectivity, but *not* the conclusion that this notion in fact applies to anything, i.e. that there are any such things as objective deontic statuses or conceptual contents. (2005, 155; footnote omitted)

The claim is that fallibilism, in the sense of structural socio-perspectivalism, cannot 'institute' *objective* normative statuses or confer their contents in the sense that such statuses or contents could be *asserted to exist* in the scorekeeping vocabulary. According to Laurier, the logical difference between the predicates 'assumes to exist' and 'exists' entails that we cannot infer the latter from the former even if it is true that discursive subjects *necessarily assume* objective normative statuses to exist.³⁸

Since Rosen (1997), many critics have raised doubts about Brandom's ideas on expressive completeness and the combination of normative phenomenalism and phenomenalism about norms. They question Brandom's explanation of (i) instituting norms through attitudes and (ii) explaining intentional concepts using normative ones. If the point is not to offer a reductive account of discursive normativity, hence meaning and intentionality, in what sense is the project 'explanatory'? Perhaps the starkest formulation of this objection comes from Sebastian Rödl (2010).

38 In his reply to Laurier, Brandom (2005, 245) notes that the difference between attributing an attitude and a status can emerge via the difference of the scorekeeper attributing and undertaking a commitment. However, as will be discussed more thoroughly in Section 11.4.2, no scorekeeper can make the distinction between what is correct and what is merely taken to be correct in his own *current* case.

Similarly to Rosen, Hattiangadi, Laurier, Glüer and Wikforss, I read Rödl as seeking to uncover a profound contradiction in MIE that sparks between phenomenalism about norms and normative phenomenalism. His more particular core claim is that, if MIE is to meet its explanatory aim of giving an adequate account of intentional and semantic phenomena by normative concepts, the normative concepts themselves must be explained by non-normative, 'naturalistic-empirical' ones. This cannot be done, argues Rödl, because the intentional and semantic concepts we actually use are 'logically dependent' on normative concepts. So, it follows that if the reductive half of MIE's aim succeeds, it does not *really* talk about *our actual* intentional and semantic concepts, and if the reduction fails, the project is not *really* 'explanatory'.

The key move Rödl makes concerns the rendering of 'logical dependency'. In his words, it means that:

[c]oncepts in a range F may be related to concepts in a range G so that one is able to use F-concepts only if one is able to use G-concepts. If we suppose that the capacity to use a concept includes the capacity to employ it in judgments, then this comes to: one is able to make F-judgments (judgments that employ F-concepts) only if one is able to make G-judgements. (Rödl 2010, 66)

The relation of logical dependency may either stem from the 'physical constitution' of the subjects in question, or else it may be 'grounded in the concepts themselves'. As an example, Rödl offers Sellars' analysis in *Empiricism and Philosophy of Mind* of 'seems'-concepts and 'is'-concepts, where the former depend logically on the latter. Be that as it may, the more relevant case of logical dependency in this context holds between normative and intentional/semantic concepts:

In what sense are elements of our mental life essentially normative? As paradigmatic of such elements, consider asserting and acting. An assertion lays itself open to assessment as true or false. It may be assessed along many dimensions; as an assertion, however, it is true or false. The dimension of correctness – call it the dimension of epistemic correctness – in which an assertion is placed as it is attributed a truth value is essential to asserting. A performance is an assertion if and only if it is correct or incorrect in this sense, i.e. if and only if it is true or false. So an act of asserting as such has a normative status, a status of epistemic correctness. Now, an assertion not only is true or false. In making an assertion one presents it as true. Thus making an assertion is adopting a normative attitude in the dimension of epistemic correctness. An assertion has a normative status and expresses a normative attitude. (Rödl 2010, 67, footnote omitted)

Once this much is granted, the fatal dilemma Rödl seeks to identify in MIE becomes clear. The normative primitives employed by Brandom, namely the concepts of discursive commitment and entitlement and the attitudes of attribution and acknowledgement, cannot hope to explain our actual intentional and semantic concepts, such as assertion, action, and perception, precisely because the intentional and semantic concepts are essentially normative in the sense of being logically dependent on normative concepts while the technical normative concepts are not; they exhibit merely 'basic-correctness'. According to Rödl, Brandom's technical concepts are definable in non-normative terms by external sanctions, which means they are not essentially normative in his sense. But if the technical concepts are not essentially normative, they cannot explain our actual normative-intentional-semantic concepts because these concepts can only be understood in terms of other normative-intentional-semantic concepts.

Rödl concludes that MIE contradicts itself by claiming that normative statuses would be instituted by attitudes and that the account is expressively complete. The external interpreter tasked to ascribe normative statuses to the performances of the pre-discursive community capable only of implicit normative attitudes and statuses may come up with translations of the performances of the form '*v* is correct'. However, while the implicit statement, which the theorist expresses as '*v* is correct', is genuinely normative for the non-discursive participants, it is not that for the interpreter, who only treats the status-description as a function of the attitude, not as an evaluative standard to which she herself would submit. So, the expression is not genuinely normative for the interpreter if she uses the technical, basic-correctness resources of MIE to ascribe it; but if she does not use those resources, she is not explaining the practice, for she has to employ the essentially normative-semantic-intentional concepts of the practice itself and become a member therein³⁹ (Rödl 2010, 74).

³⁹ In his response to Rödl, Brandom (2010a) denies that the scorekeeping model would be reductive of normative concepts since it holds onto normative phenomenalism. This reply is interesting in that Brandom seems to be implying that whatever *explanatory* virtues the scorekeeping model may have would hinge on its reductive capacities, although he also says not to 'systematically reflect the distinction' in his rhetoric.

10.2 Against Normative Phenomenalism

Lionel Shapiro (2004) observes a yet distinct contradiction in the great programmatic synthesis between phenomenalism about norms and normative phenomenalism. His main idea can be summarised as follows. First, Shapiro notes that MIE purports to explain (a) in virtue of what our assertion tokens and other linguistic expressions come to have the meanings they do, and (b) what is it that the meaning consists of. Second, Brandom is seen to deploy a broadly pragmatist-expressivist strategy to bear on both questions; he tries to explain the correctness of meaning ascriptions and attributions by the normative practice of scorekeeping. The meanings of our expressions are determined by the normative practice. On the second (b)-question, Brandom, however, answers by claiming that it is not scorekeeping as *actually* done by the scorekeepers of which meaning consists, but in how the score *ought* to be kept. Third, Shapiro attempts to show that due to Brandom's commitment to deflationism about truth, he is neither required nor entitled to give *any* answer to the constitutive (b)-question.

Let us go over Shapiro's arguments in more detail. In programmatic terms, phenomenalism about norms strives to explain the meanings of token-utterances, most importantly, assertions, by what it is for subjects to take or treat those utterances to have meanings. This is the expressive-pragmatist reductive half of the project. The key concept here is the attitude of attributing a discursive commitment, which in turn is explained in the expressive-pragmatist fashion by studying what an external interpreter *does* in attributing the attitude of attributing a discursive commitment. In this sense, says Shapiro, the strategy is 'normative *attitudes* all the way down' (2004, 142). Shapiro's ultimate point is that because Brandom is a deflationist about truth, he is, by that token, not required to say anything more about ascriptions of meaning other than what his scorekeeping account says about making explicit the implicit attributions of meaning. In particular, following Horwich's example, it is not a criterion of adequacy for a semantic theory that posits some property or feature 'in virtue of which' an arbitrary sentence means what it does that

the *truth conditions* could be derived from the said property/feature.⁴⁰ 'In practice', this is exactly what Brandom does, according to Shapiro's reading (2004, 147-148).

The problem is that Brandom explicitly denies what he allegedly does in practice, or at least what he, according to Shapiro, should do. Namely, under the programmatic call of normative phenomenalism, Brandom purports to give a *separate, non-deflationary* answer to the 'What meaning consists of' question than what his pragmatist method delivers in answer to the determination question. The goal here is to arrive at a theory of objective meanings and conceptual content such that it is, in Crispin Wright's terms, 'ratification independent' (or in Brandom's own, 'attitude-transcendent'). The main consequence of attitude-transcendence is that, according to Brandom, scorekeepers can be committed to claims, which no one in the practice has attributed (or even ever will attribute) to them, nor which they acknowledge themselves. This is precisely the bone of the contention that the attitudes 'institute' normative statuses, which in turn confer the contents, or in which sense it is '*statuses* all the way down' after all (Shapiro 2004, 143).

Shapiro argues that Brandom (a) has not delivered a substantial account of the institution metaphor, and (b) he neither needs nor can deliver it with the resources he affords himself in MIE. How is it exactly that Brandom is *precluded* from giving such an account? The reason is that it follows from rejecting the derivability constraint that there just is, in Brandom's own terms, *no pragmatic work to be done* for the normative statuses understood in the attitude-transcendent sense. With the rejection of the fundamental *reason* to meet the derivability constraints also goes the *possibility* of meeting it (Shapiro 2004, 157).

Shapiro's idea that once we take phenomenalism about norms seriously, there just is not anything left for normative phenomenalism to explain, is echoed by Ronald Loeffler (2005). The main difference in terms of the outcome is that whereas Shapiro shows how phenomenalism can and should be made consistent with its inherent semantic deflationism, Loeffler radicalises it further by arguing for full-blown meaning eliminativism.

⁴⁰ More accurately, the 'derivability constraint', as Shapiro calls it, is trivially satisfied by the theory by the assumption that the key feature/property just is in virtue of which an arbitrary sentence has the meaning it does.

How big is the gap between deflationism and eliminativism? The issue is delicate, for both isms can be formulated in a number of non-fungible ways, but since the synthesis I am examining gravitates on programmatic orbits, it suffices to give a cursory explanation of the difference here. For one, we must distinguish between deflationism about meaning (propositional content, including beliefs and other propositional attitudes) and deflationism about externalist semantic notions, such as truth, reference, and representation. Brandom is explicitly a deflationist about the latter, but he is *not* a self-avowed deflationist about meaning or propositional content. In other words, there is a substantial philosophical story to be told about meanings and propositional attitudes, although not one in which meanings or attitudes are understood *solely* as properties of subjects. As we recall from Section 9.4, the algorithmic dispositions are a necessary condition for propositionally contentful attitudes, but not a sufficient one.

The problem is that for Brandom to hold a non-deflationist account about propositional contents, he must tell a story in which the contents are determined not just by the attitudes but by the world. But since he is deflationist about the usual externalist explanatory candidates, he must explain how the constraint emerges from the side of the attitudes and not from the world directly. This is the task he meets under the heading of normative phenomenalism. Shapiro's ultimate point that we already discussed is that this cannot be done; as a result, deflationism about the representational predicates entails deflationism about propositional contents, against Brandom's explicit aims. If there is no story to be told about how propositional contents can be determined by the world (and not just by the attitudes), the account becomes semantically deflationary in both intensional and extensional sense in Shapiro's view. The main difference to Loeffler's eliminativism is that whereas deflationism states there is no substantial philosophical explanation to be given about propositional contents, eliminativism states that such contents are to be 'explained away'. The first is a merely negative thesis about the limits of philosophical explanation, the other a positive metaphysical thesis.⁴¹

⁴¹ Sebastian Knell (2005) also *suspects*, though does not argue, that Brandom is driven to deflationism about propositional contents due to deflationism about representational predicates.

10.3 Distinguishing between Normative Force and Content

We have now reviewed (examples of) both branches of the criticism, which Brandom's bold synthesis between phenomenalism about norms and normative phenomenalism faces. The first branch argues that the synthesis fails from bottom-up, that there is no successful explanation of how discursive norms can arise from the pre-discursive base practices and abilities that does not already presuppose intentional concepts in some form, either in the subjects themselves or by their external interpreters. As a result, genuine normativity remains unexplained, although not inexplicable, and though the full ambitions of the scorekeeping model cannot be realised, the account is undeniably philosophically valuable in other ways.

The second branch rather sees that the problems are not with the phenomenalist-expressive-pragmatist base explanans, but with the higher reaches of the explanandum. The scorekeeping model cannot explain 'meanings' or 'norms' 'in themselves', for all the reality of meanings and norms can be understood fully in terms of the attitudes and their interaction. In other words, phenomenalism about norms has neither the possibility nor need to meet the requirements of normative phenomenalism.

The red line in all these arguments has been, I claim, a concern with the *objectivity* of norms. This should not be very surprising considering the grand architecture of MIE, which aims to account for objectivity in terms of subjective practices and abilities. However, I think that the preceding literature has not always done well to distinguish between two importantly different categories of objectivity, namely the *pragmatist* and the *semantic* varieties. The *locus classicus* of objectivity has been interpreted in the semantic vein, namely by how Brandom attempts to explain how discursive norms implicitly instituted by practices may confer propositional contents on assertions and inferences. The reality of the contents is then directly dependent on their objectivity, i.e. their power to represent worldly objects, while the reality of norms themselves depends directly on their contents. Whether the explanatory order from institution to conferral can be made to work in a sufficiently strong sense of semantic objectivity has been the major discussion point in the debate. In contrast, the question regarding the objectivity of the institution of implicit norms themselves has not been so extensively examined. Indeed, the very idea of such 'pragmatic

objectivity' of the norms themselves may initially appear like a misnomer, for surely no one thinks that the normative attitudes as *acts* are supposed to represent anything – that is the job of their contents after all.

In the following, I shall argue that there is an interesting sense of 'pragmatic objectivity' inherent in Brandom's works, which is intricately connected to, yet distinct from, semantic objectivity. Briefly, whereas the semantic objectivity of norms concerns their (primarily propositional) contents, the pragmatic objectivity of norms concerns their (primarily obligating) *force*. By 'normative force', I mean the sense in which any (discursive) norm is or becomes *binding* on subjects, and particularly in what sense various acts of bindingness can be said to be objective.

As I shall argue in Section 12, perhaps the greatest challenge to Brandom's bold synthesis comes from the specific angle of the pragmatic objectivity of normative force and how it can become genuinely binding on subjects. The reason it is the greater challenge is that, as I shall argue in Section 11, Brandom's post-MIE developments, in fact, enable him to offer a more robust sense of semantic objectivity than his early critics have thought possible, although at the cost of significantly expanding the project launched in MIE. To anticipate, the way the scorekeeping model can account for a strong notion of representational purport is by joining forces with 'conceptual realism', which is the idea that the non-discursive world itself is conceptually structured.

11 SEMANTIC OBJECTIVITY

This section examines Brandom's conception of semantic objectivity and certain criticism that has been posed to it. My overall goal here is to defend Brandom, for I believe that already in MIE, he can provide a *principled* defence for the objectivity criterion that he sets on himself. However, the road is not as straightforward as he claims it to be, and it takes some reconstructing to make work. More particularly, from my perspective, Brandom is not as explicit as he should be about the role that *conceptual realism* plays in his semantic objectivity account in MIE. I argue that conceptual realism (i.e. the metaphysical claim that the world itself is conceptually structured) plays a central part in making the objectivity account work as intended, although in MIE, Brandom does not provide independent support for the thesis.

The section will proceed as follows. In Section 11.1, I will distinguish between three levels of semantic objectivity for norms, which helps to understand some of Brandom's major claims and goals in MIE. My main argument here, through Sections 11.3, 11.5, and 11.6, is that there is a discrepancy between Brandom's official and less official characterisations of the standard for semantic objectivity that he sets for the scorekeeping account in MIE. The discrepancy has to do with the sense in which the normative practice can be said to represent the world of objects by 'incorporating' them in the practice. Furthermore, I argue that Brandom also has a principled way to bridge the discrepancy, which is the idea of 'conceptual realism', or that the world itself is somehow conceptually structured and thus fit to be incorporated in the practice. So, conceptual realism is the ultimate guarantee of semantic objectivity in MIE, and semantic objectivity thus amounts to something more than a structural feature of the practice. (To anticipate, in Section 12, from my standpoint, I will argue why Brandom indeed needs the more demanding standard of semantic objectivity – to support his commitment to the grand synthesis and the genuineness of norms implicit in practice.)

Aside from this main line of argument, I also defend some aspects of Brandom's semantic objectivity account against specific arguments by Thomas Fossen (2014) (Section 11.2.) and Sven Rosenkranz (2001; 2003) (Section 11.4.1).

11.1 Three Grades of Semantic Objectivity

At the lowest tier of objectivity for norms, we find *attitude-immanent* norms, which can be characterised as follows:

(AI) A norm n is attitude-immanent for community C iff it is not possible for everyone in C to be mistaken about the content of n .

Prime examples of (AI)-norms are social norms, such as greeting gestures and marriage institutions, in case of which 'it makes no sense to suppose that [the community] could be wrong about this sort of thing' (MIE, 52-53). A few remarks should follow the biconditional definition. To be wrong about the content of a norm means that not everyone in the community could be incorrect in their practical evaluations of what counts as correct and incorrect behaviour according to the norm, nor which kind of behaviour should be counted as falling in its purview to begin with (i.e. what kind of behaviour is said to be *relevant* to the norm). Moreover, the principle is bound to incorporate an ineliminable measure of vagueness regarding how finely the norm's content should be individuated, for it is typically the case that community members do not have robust evaluative intuitions or dispositions about all possible circumstances in which the norm could be applied. So, I take it to be compatible with the (AI) status that a norm's content is not wholly transparent to the community in the positive sense that they could not find genuinely novel, yet unthought of, circumstances of application for the norm, although they could not then all be incorrect about the application of the norm. Finally, the collective judgement can be represented either by all the mature members separately or by some select, deferrable group of experts among them.⁴²

⁴² It should be remembered that even in cases where the community members defer to, e.g. sauna experts, to decide what the proper sauna norms are, the account still counts as *I-thou* as opposed to *I-we* in kind for Brandom, for the position of certain individuals as experts is itself a normative matter that is ultimately decided by the whole community's recognition.

The class of norms that the objectivity of which (AI) grading most readily befits are often called 'social norms'; a slightly misleading term since all norms have a social character in some sense, at least for Brandom. While much and more could (and should) be said about attitude-immanent social norms, e.g. how to distinguish them from mere conventions⁴³, the important point here is to contrast them with attitude-transcendent norms:

(AT) A norm *n* is attitude-transcendent for community *C* iff it is possible for everyone in *C* to be mistaken about the content of *n*.

In contrast to (AI), a norm that is (AT) has contents, which are not necessarily and sufficiently determined by the community's collective judgement. In particular, Brandom argues that we must understand conceptual norms as distinct from merely social ones precisely in that only conceptual norms are rightly called attitude-transcendent: it is a condition of adequacy he sets for his own account that the norms constituting the meaning of expressions (or more generally, conceptual contents) should count as attitude-transcendent (MIE, 53-54).

Whether the aspiration succeeds is, of course, another matter entirely. To begin to see the challenges Brandom faces here, it is important to note that attitude-transcendence is not yet sufficient for what I shall here call 'properly objective' semantic objectivity grading for norms as depicted by:

(PO) A norm *n* is properly objective for community *C* iff the world of objects determines the content of *n*

where 'the world' is understood encompassingly as whatever exists (metaphysically) independently of discursive practices. The crucial difference between (AT) and (PO), noticed especially by Loeffler (2005, 57) and Prien (2010), is that the former stands only for the negative claim that a norm's content is not determined by the community's attitudes, while the latter states the positive thesis that it is the world that determines the content.⁴⁴ Strictly speaking, for all that (AT) says, it is compatible

⁴³ I refer the reader to Brennan et al. (2013) for a thorough conceptual study on social norms.

⁴⁴ This is not to say anything about the degree to which the world determines the contents of the commitments that are about it. An extreme realist view might hold that there was only one correct

with it that the content of the norm *n* is indeterminate, i.e. determined by nothing, or then determined by a Cartesian demon.⁴⁵ Brandom's realist challenge in the semantic dimension of objectivity is, then, not only to explain how norms correctly depicted by (AI) can give rise to norms (AT), but also what more is required for (AT) to yield norms in the sense of (PO).

There are two important points that can be made based on the simple objectivity hierarchy for norms. First, Brandom's basic, major claim in MIE is that (AT) norms can be understood as arising from the more basic (AI) norms, which amounts to the claim that non-conceptual, non-discursive normative practices can institute conceptual, discursive ones, i.e. the programmatic claim of phenomenalism about norms. While (AI) norms can be understood as basing their objectivity solely on the sum of attitudes in the community, (AT) norms somehow are supposed to transcend the attitudes in terms of semantic objectivity.

The second important point is that in MIE, Brandom apparently sets himself the more demanding standard of (PO), yet what he can show holds true in the scorekeeping system is only the (AT) standard. So, there is a gap left between the claims that (i) the attitudes do not determine what is correct in the scorekeeping practice and (ii) the world has (extended) authority to determine what is correct. In Sections 11.5 and 11.6 below, I will argue that it is conceptual realism, which is supposed to bridge this gap.

Another way to describe the difference between (AT) and (PO) norms is that only the latter affords the concept of representational purport, the explanation of which Brandom takes to be an important evaluative criterion for his project. The reason (AT) norms arguably do not suffice for representational purport is that, as was seen above, a norm being (AT) does not exclude the possibility that its content is indeterminate (i.e. determined by nothing) or then it is determined by a Cartesian demon. Were that the case, it becomes hard to argue that the norm purported to

way to represent the world, in its parts or its entirety, with no room left for alternative schemes. Aside from deeming such a view intuitively unconvincing, I leave the question over degree of determination open here.

⁴⁵ Technically speaking, the Cartesian demon is also assumedly part of the world, and so this possibility would be compatible with (PO). The point is though that the demon could determine the norm's content independently of what is actually the case.

represent *anything*, much less the world, which is why the claim to (PO) status for conceptual norms is crucial for Brandom to achieve.

11.2 A complication regarding the objectivity grades

Before moving on to discuss how Brandom's project for properly objective conceptual norms might be plausibly realised, I want to draw attention to a possible complication that I think has not been adequately discussed in the literature regarding the relations between the three grades of semantic objectivity for norms. The problem is that although (AI), (AT), and (PO) have sufficiently clear definitions in the abstract, it seems to be less clear how the community itself can distinguish between particular norms as falling under any of them. The point is appreciated by Thomas Fossen, who writes that:

the idea that a community cannot intelligibly be taken to be wrong about who appropriately counts as married seems to be invalidated by controversy over gay marriage. Brandom's use of these examples only works on the assumption of a mythically homogeneous community where there is no disagreement on these issues. (Fossen 2014, 387)

I think Fossen taps into an interesting tension in Brandom, although the line he takes faces some difficulties that I shall briefly elaborate here, with an intention not so much to solve as to highlight them for the literature.

In criticism of Fossen's remark above, from the fact that, e.g. marriage norms, are actually contested in some communities, it does not follow that marriage norms would not therefore count as attitude-immanent; it only follows that there is disagreement among the members on what marriage norms *should* be accepted by everyone. *If* everyone in a community were to agree on what counts as a genuine marriage, *then* the content of the norms would be determined, which is compatible with the observation that not all communities have collectively accepted marriage norms in this sense. So, unlike what Fossen seems to imply, the actual fact of controversy over marriage norms does not entail that a *whole community* could be intelligibly taken to be wrong about their contents.⁴⁶

⁴⁶ Implied here is the deep, difficult question of how we should understand community-membership. On the face of it, if there is widespread disagreement among a group of people about what marriage

The more pressing question, I believe, is whether the difference between attitude-immanent, attitude-transcendent, and properly objective norms is itself attitude-immanent, attitude-transcendent, or properly objective. In other words, is the community's judgement necessary and sufficient to determine whether a norm is (AI), (AT), or possibly (PO)? Fossen raises a similar question against the intuitive idea that, while some norms, e.g. that raising one's right hand counts as a greeting, are purely social and (AI), other, conceptual norms are objective, and (PO), for example, the commitment to the claim that 'The earth is flat'. However, he observes that:

it does not follow that conventions are therefore uncontestable, and a community is incorrigible about them. The key is to see that whether something is arbitrary, and therefore fully settled by mere convention, or objective, settled by the way the world is, is a normative question. It depends on a background of further commitments. Whether or not the earth is flat is not arbitrary in this sense due to a commitment to getting things right about the world (implicit, as Brandom argues, in discursive practice). Whether it is appropriate to raise one's right or one's left hand as a greeting is arbitrary only if we take it that the community is not already committed on the matter in virtue of some other consideration. To take something to be a matter about which the community cannot be mistaken is to take it that there are no incompatible commitments on the matter, and hence to attribute an entitlement to the community to settle it in one way or an other. (Fossen 2014, 387-388)

I agree with the notion that a norm may be attitude-immanent and still contested within the community, but I see a problem in Fossen's claim that the difference between (AI) and (PO) norms would be 'normative', which he himself does not seem to acknowledge or at least discuss, although I believe that this reading meshes well with Brandom's overall program. For one, the characterisation of the difference as normative leaves open to what grade of semantic objectivity the normativity aspires. The two main options both come with problems, as I shall now argue. The starting point is that if the question of what objectivity grade a norm belongs to is itself a

norms should be accepted, what would settle the question of whether the controversy sparks *within* a single community or *between* two different ones? Brandom's general answer in MIE (39-40) is that delineating who, in fact, belongs to a community is a normative matter as opposed to a causal one. This claim is compatible with multiple different readings of 'normative', especially considering the three tiers of objectivity mentioned above: there can be communities *qua* discursive practices and *qua* marriage norms, and the two need not conflict. In this instance, I cannot address the matter any further.

normative question, then there must be a further norm that determines this; hence, we can ask how objective the content of this further 'metanorm' is.⁴⁷

To begin with, assume that the difference between a norm being (AI) versus (PO) is itself objective in the (AI) sense. It follows that the community cannot err about what its collective judgement, if there is any, would be about which norms count as attitude-immanent, therefore settled by the attitudes only, and which count as properly objective, settled by the world only. Under this assumption, it would be coherent to speak of a deeply religious community in which there was universal agreement that their marriage norms are properly objective, settled by their deities as a matter of (super)natural fact, as *determining* that their marriage norms were indeed (PO). The marriage norms would then have representational purport for the community members simply because they collectively judged the matter to be so. Call this possibility the (AI) reading.

Of course, although it is plausibly up to the community's judgement whether a norm of theirs has representational purport, it cannot be similarly up to their attitudes alone whether the norm *succeeds* in representing anything in the world, simply because the attitudes alone cannot determine what there is in the world (in this case, whether the deities exist or not, or whether they have been correctly interpreted). Nonetheless, it might be maintained that no community can have the power to determine even whether one of their norms has representational purport or not. The suggestion would then be that whether a norm is (AI) or (PO) is determined by the world in the (PO) sense, hence that the religious community, regardless of what their collective views are on the matter, was simply wrong about whether their marriage norms had representational purport. Only if there was something in the world (e.g. divine commands) to *be* represented by the norms could the norms so much as *aspire* to have representational purport; the (empirical) *potential* for representational success is a necessary condition for representational purport on this reading⁴⁸, and since

47 Some might worry that if determining which semantic objectivity grade a given norm belongs to is itself a normative matter (hence open to interpretation), a regress of interpretations is inevitable since presumably, the metanorm(s) must be also interpreted in order to yield any results. But the regress only follows if we require the interpretation to always pivot on an explicit norm as opposed to an implicit one, which is the fallacy of 'regulism' (MIE, 20).

48 For example, an error theorist would undoubtedly deny this; here, I am not claiming that she would be wrong to do so, although I have my doubts. For I do not think it is obviously coherent to talk of

representational success is not even potentially up to the community attitudes, neither is the purport. Call this possibility the (PO) reading.

The problem with the (PO) reading is that communities could, according to the (PO) reading, be in principle intelligibly wrong about the objectivity status of their own norms. While this makes intuitive sense in cases, such as the one described above where a norm taken by the community to be (PO) is actually (AI), the converse possible case where a norm universally taken as (AI) is in fact (PO) seems more difficult to imagine. For instance, assuming a community that unanimously accepted that raising one's right hand counted as a greeting, and who also accepted that they could not be intelligibly wrong about this, the possibility would remain under the (PO) reading that the community was indeed wrong about the objectivity status of their norm, the content of which is actually determined by the world. But what else, save a Cartesian demon, could possibly invalidate their attitudes on the metaquestion?

Strictly speaking, both (AI) and (PO) readings are logically coherent; their respective problems that I have briefly characterised are only *prima facie* in nature. Nonetheless, I believe the questions represent interesting issues for further discussion, while also revealing some characteristic aspects of Brandom's general account about objectivity. I shall further examine these general facets of objectivity in the next subsection.

11.3 Brandom's Objectivity: Determination as a Negative Concept

One comparatively clean resolution of the problematic examples depicted above would be to point to a confusion between the epistemic and ontological senses of objectivity. The *primary* sense of semantic objectivity, we would then like to say, is ontological in that whatever objectivity grade a norm actually fits is independent from the community's attitudes, which the secondary, epistemic sense of objectivity can at most reflect but not determine. Hence, it is misplaced to ask whether the

empirical potentialities, which, as a matter of fact, could never be actualised. To say that the potentialities could be metaphysically actualised only makes things worse, I feel. As was mentioned though, in this section, my goal is only to motivate the problematique.

relation between the norms of differing semantic objectivity grades could be understood based on the grading itself. This suggestion, I believe, amounts to the negation of Fossen's claim: the objectivity grading of norms is not itself a normative matter.

However, this neat rendering is not available for Brandom, for he cannot privilege the ontological sense of objectivity in case of norms because of his methodological commitment to phenomenalism about norms, which states that *all* (at least broadly discursive) norms are instituted by our attitudes. Hence, we cannot derive the objectivity status of the norm from the norm itself considered as an entity existing independently of the attitudes, but we must construe its objectivity status from the community attitudes. But this methodological primacy of the epistemic sense of objectivity precisely then leads us to ask in what sense a community's attitudes are to determine the objectivity grade of their own norms.

We are now in a position to note another general, important feature about Brandom's approach to objectivity aside from the privileging of the epistemic sense. The dimension on which the three grades of objectivity align is defined, not by a kind of *a positive* feature, such as correspondence to reality or internal coherence, but an essentially *negative* trait, according to which the most objective commitments are those which can be found to be 'most' erroneous. At the (AI) tier, error is possible only for individual subjects but not for the whole community, while at the (PO) tier, whole communities may be found to be in error regarding their shared commitments. This result is a fairly straightforward consequence of Brandom's methodological choice to privilege the epistemic sense of objectivity in case of norms. The shape of his challenge is not to explain how attitudes could *correctly* get a grasp of conceptual norms considered to exist independently from our attitudes in some sense, but to show how, starting from what individual subjects *take* to be correct in practice, they can come to 'institute' norms objective enough, so that they can be *incorrect* in their takings by their 'own lights'.

Furthermore, Brandom is a stout fallibilist about conceptual contents in the sense that in his view, there are no justifications or warrants strong enough to *guarantee* the correctness of certain conceptual contents, but because of conceptual contents' *essentially perspectival nature*, i.e. the fact that the contents can only be specified against

a background of auxiliary concomitant commitments, which vary from subject to subject, all conceptual commitments are perpetually liable to be found incorrect from some further, perhaps never-to-be actualised discursive perspective, all formally equal with each other in this respect (cf. Section 9.2 above).

Above, I claimed that the attitude-transcendence of a norm is not sufficient for it to have representational purport. I now claim that the same point applies to the more general fallibilist thesis, which generically states that our very best warrants for a claim may always turn to diverge from the truth: fallibilism thus understood is not sufficient for representational purport. In order to represent anything beyond their practice, the *source* of the incorrectness of claims made in the practice must belong to what is being represented, namely the worldly objects. As a species of fallibilism, (AT) is not sufficient for representational purport because it does not specify *what does* determine which commitments are incorrect. We cannot assume that since attitudes do not determine correctness, and because the subjects take it that their attitudes do not determine correctness, it is the world that determines correctness, or is even taken to determine. For the source might as well in fact be a Cartesian demon; and it would be queer to say that the practice then purports to represent the demon.⁴⁹

At this point, one might be struck by an exegetical worry: is it really true that Brandom is committed to reaching (PO) as a standard of semantic objectivity in MIE, as opposed to being satisfied with (AT)? While it is true that the official “objectivity proofs” (MIE, 601-607; see below in Section 11.4.2) only rely on the (AT) standard, Brandom writes about semantic objectivity in a more demanding sense at several places:

The objectivity of conceptual norms requires that any attitude of taking, treating, or assessing as correct an application of a concept in forming a belief or making a claim be coherently conceivable as mistaken, *because of how things are with the objects the belief or claim is about.* (MIE, 63, my italics)

The *objectivity* of representational content is a feature of the practices of assessing the correctness of representations. The status of representings as correct or incorrect, successful or unsuccessful, depends on how things are with what is represented, rather than on the attitudes of representers. What is distinctive of specifically

⁴⁹ If this is correct and fallibilism does not entail representational purport, it becomes doubtful whether the thesis counts as a realist stance at all, unlike how it is usually interpreted (see, e.g. Rydenfelt 2019).

representational correctness is this objectivity—the way in which assessments of *representational* correctness take representings to answer to what is represented, rather than to how what is represented is taken to be. It is the way in which the status being assessed outruns any particular attitude toward it. Understanding the objectivity of representational content requires understanding this particular structure of its authority and acknowledgement—what it is for those assessing the correctness of representings to cede authority over them to what is represented, to treat their correctness in practice as determined by those representeds. (MIE, 78)

There is another exegetical objection one might make against the idea that (PO) is Brandom's ultimate semantic standard of choice in MIE. In discussing the notion of semantic objectivity in MIE, and how the normative statuses instituted by attitudes can outrun the attitudes in terms of what is really correct (MIE, 631), Brandom mentions a brief analogy to games. The scorekeeping practice is 'solid' or 'lumpy' in a way baseball is, and in a way unlike chess is, in that it necessarily incorporates physical objects with physical properties in its language entry and exit transitions. Here, concrete worldly objects are involved or incorporated in the practice in the (AT) sense because it is possible that a whole community is wrong, e.g. about the real weight of a given baseball (MIE, 632, fn.).

Notice, however, that the baseball itself does not have any authority as to what its *correct* properties, e.g. weight, are, for these are set by the attitudes alone. A given community might err in measuring the weight of any ball, but it cannot collectively err about what the correct weight is supposed to be. In contrast, Brandom is clear that the whole community can err about what the correct melting point of copper is, and that this is central in the community's ability to represent copper with terms like 'copper' *because it is copper itself that determines its correct melting point*. So, there is actually a stark disanalogy to the 'lumpiness' of baseball and the scorekeeping practice in that the latter is meant to attribute authority to the very objects it seeks to represent. This, I maintain, just is the (PO) standard, which Brandom takes on in MIE, albeit not as explicitly as he perhaps should.

11.4 External Criticism of Brandom's Objectivity

The external criticism of Brandom, as discussed here, stems primarily from his ambitious aim to combine two claims. The first is that assertions are primarily

knowledge claims (MIE, 227). The second is that all three key elements of the traditional concept of knowledge, namely justification, propositional content, and truth, can be explained in terms of a certain normative social practice that he calls 'deontic scorekeeping'. In the following subsection (11.4.1), I shall focus on Sven Rosenkranz's criticism that the scorekeeping practice cannot explain our ordinary understanding of truth in relation to its warrants. In Section 11.4.2, I focus on another facet of our so-called ordinary notion regarding truth and belief, and how well the scorekeeping account can pay homage to it.

11.4.1 Warrants

For one, Rosenkranz (2001; 2003) has argued that incompatibility semantics are incompatible with some aspects of our (allegedly) pre-theoretic understanding of truth and warrants for it, to which he reads Brandom at least as trying to pay homage. In particular, Rosenkranz appeals to the idea that it is (necessarily) coherent to claim that a commitment (or a belief) may be objectively true, although we have no epistemic warrants to it. However, he argues that Brandom's fundamental concepts of discursive commitment and entitlement conflict with this principle:

Since what one is committed to is a function of what others are authorised to treat one as being committed to (MIE, xiv, 142, 161-163), whether one acknowledges these commitments or not (MIE 194-195), the commitment to a given claim has as its pragmatic consequence the commitment to the existence of warrants for this claim. (Rosenkranz 2001, 233)

According to Brandom, in acknowledging a commitment to a claim (paradigmatically, by making an assertion), the subject undertakes a conditional task-responsibility to justify the commitment if appropriately challenged to (MIE, 172). According to Rosenkranz, this means that commitment to p pragmatically entails commitment to there being entitlements for p . But this is clearly in contradiction with the principle that it is coherent to claim that there are truths for which we do not have, and perhaps could never have, justifications for, perhaps the most famous example of which being Golbach's conjecture.

Now, Brandom himself has emphasised that the commitment to p should *not* be identified with the commitment to being entitled to p : 'For surely one could be

committed to [p], without thereby being committed to the claim that one is *entitled* to [p]’ (AR, 200). Rosenkranz observes that this seems to conflict with Brandom’s earlier statements both in MIE and earlier, such as:

In asserting a claim one not only authorises further assertions, but commits oneself to vindicate the original claim, showing that one is entitled to make it. (1983, 641)

In his analysis of this tension, Rosenkranz writes:

the fact that one may be committed to p without being entitled to this commitment does not show that one may be committed to p without thereby being committed to the claim that one is so entitled. (2001, 233)

In my view, Rosenkranz holds a mistaken diagnosis of Brandom’s account of commitment. The mistake is to identify Brandom’s claim that:

- (1) commitment to a claim pragmatically entails a conditional task-responsibility to show an entitlement to the claim

with the claim that

- (2) there are entitlements to the commitments one acknowledges or that one is able to show them.

That one is subject to the conditional task-responsibility to show entitlement to p if appropriately challenged to, which is a pragmatic consequence of acknowledging or undertaking the commitment to p (MIE, 172), does not entail the claim that one is committed to the claim that there *are* entitlements or warrants for p . For the existence of the responsibility does not entail, pragmatically or otherwise, that it can be fulfilled, or even that the subject is committed to claim the possibility that it can be fulfilled. True, this does violate the widely cherished 'ought implies can' principle that is usually accepted in the case of specifically moral norms. But nowhere in MIE does Brandom claim that the principle applies to discursive norms, and elsewhere, he explicitly asserts that:

we may not *actually* do what is in this [discursive] normative sense demanded of us, or even practically be *able* to do it. We are discursively born into a state of sin, for all our conscientious efforts, are by and large doomed to live in such a state.' (BSD, 191)

The distinction between (1) and (2) shows that acknowledging the commitment to p need not be incompatible with the commitment to the claim that 'I currently have no warrants to p '. Although Brandom thinks that such 'bare assertions' (MIE, 229) must be understood as parasitical on proper assertions that shoulder their responsibility to show entitlement to the commitment if appropriately challenged to, the potential to make bare assertions in the scorekeeping practice suffices to save the distinction between claiming p and claiming to have warrants to p , which Rosenkranz argues is lost by Brandom.

I believe the same analysis can reject Rosenkranz's slightly different formulation of the same point:

When committing oneself to A one cannot, at the same time, coherently commit oneself to anything inconsistent with the claim that there are presently available warrants for asserting A , i.e., WA . For, if one commits oneself to something inconsistent with WA one commits oneself to $\sim WA$, and if one commits oneself to $\sim WA$ one cannot coherently seek to present oneself as someone who qualifies as knowing A , and that's what one seeks to do when committing oneself to A . (Rosenkranz 2003, 343)

If this much is correct, then material incompatibility relations cannot explain what it is for two claims to be incoherent. For although all incompatible claims will be incoherent, not all incoherent claims will be incompatible. For if A and $\sim WA$ are incompatible, i.e. commitment to one precludes entitlement to the other and vice versa, then it should be incoherent to claim that A is true and that there are currently no warrants for A . But this is clearly not incoherent, e.g. with Golbach's conjecture: hence, Brandom's pragmatic strategy to explain incoherence fails.

Except it does not, at least for the grounds Rosenkranz proposes. The reason is that Rosenkranz misses Brandom's caveat for bare assertions, i.e. assertions for which one explicitly declines to fulfil the conditional task-responsibility to show one's entitlement to the commitment. The crucial point is that it is not incoherent to make such assertions because it is not incoherent to be subject to an obligation that one cannot fulfil, or which one knows one is not (currently) able to fulfil. True, not *all*

assertions could intelligibly count as bare assertions, which is precisely what Brandom underlines by characterising bare assertions as parasitic on proper assertions. We always ought to be entitled to our commitments, and we often are, yet we might not always live up to the obligation, which does not mean our assertions stop being assertions. Golbach's conjecture is one prime example of a claim one may assert, while also denying that one is entitled to its truth: it is thus a bare assertion, but not an incoherence.

11.4.2 Truth

How about the other half of Brandom's contention that assertions are knowledge claims, namely that every assertion *primarily* aims to be not only warranted but also true? It seems to be an important aspect of our ordinary concept of belief that I may consistently believe that some of my beliefs are false; at the very least, it seems necessary to admit the possibility of some of my beliefs diverging from truth. The question is whether the scorekeeping account can explain this facet of our ordinary concept of belief (or what analytic philosophers tend to take as ordinary at any rate), which is an important element of their objectivity.

Here, things are more complicated because of Brandom's theory of conceptual content as *essentially perspectival* in nature. As we recall, this means that conceptual contents (paradigmatically propositional ones) can only be specified in the context of auxiliary commitments that vary from subject to subject. What is shared by all subjects is that they take there to be a distinction between what anyone else takes to be semantically correct and what really is semantically correct, where the 'really correct' set is said to be those commitments the scorekeeper herself undertakes, as opposed to the sets she attributes to others (MIE, 595, 597). The problem now is: how is it possible for the scorekeeper to apply this distinction in her own case, i.e. to acknowledge a difference between the commitments she undertakes and those commitments that are really correct?

In a somewhat strong sense, the scorekeeping account cannot, I think, incorporate the idea that some of my beliefs could be false. For according to the perspectival nature of conceptual contents, two claims

(a) I claim that p

and

(b) not- p

will always be incompatible with each other (I cannot be entitled to them both), although not incoherent in the logical sense that gives rise to Moore's paradox. However, Brandom is, in my view, correct in claiming that ascribing

(c) S claims that I do not claim that p , and S claims that p

to S is not to attribute incompatible commitments (MIE, 605). Moreover, the scorekeeper can herself be entitled to both conjuncts, and in that sense, recognise the possibility that one of their own commitments is actually false *by someone else's score*. Admittedly, this is not the same as directly recognising the potential falsehood of a presently acknowledged commitment, but it suffices to falsify the threatening inference from 'I claim that p ' to ' p ', which would make first-person error impossible and hence lose a robust sense of objectivity for beliefs. For even though the 'commitment conditions' of these claims are identical, i.e. one is committed to one whenever one is committed to the other, the two diverge in terms of their entitlement conditions. So, the inference from 'I claim that p ' to ' p ' will always be commitment-preserving but never entitlement-preserving, which is reasonably taken as a minimal threshold for any successful objectivity account. This is the kernel of Brandom's objectivity proofs, or the official standard of semantic objectivity which the scorekeeping account is tasked to meet. But as we already saw above, it is not the final standard: (PO) must also be respected.

11.5 The Pragmatist Route to Semantic Objectivity of Contents Goes Via Conceptual Realism

A deep orienting commitment of Brandom is that *justification* is essentially a normative, not a causal, concept. Although he rejects the internalist claim that to have justifications for one's commitments always entails *having* or being able to *show* justifications for those commitments, he criticises the reliabilist theories of justification for being insufficient to explain the concept of justification (AR, 97). The background for this major clash in contemporary epistemology rests with Sellars' celebrated busting of the Myth of the Given, as well as with Davidson's (1986a) arguments to the effect that only belief can justify other beliefs. The common idea shared by Brandom is that merely causal relations between the world and doxastic, discursive processes are not sufficient to account for proper justifications. However, he differs from the anti-realist conclusions, drawn, e.g. by Michael Dummett and Richard Rorty, according to which the web of belief (or network of vocabularies) is nowhere 'anchored' in the world (PP, 124-127; Brandom 2019, 17). One major strategy to substantialise this anchoring is by explaining how some norms in the practice come to exhibit representational purport and success.

As I already explained, the shape of the problem of semantic objectivity for Brandom is to explain how objects of the world can come to be incorporated in or mediated by our discursive attitudes in the sense that the original authority of the attitudes is, in a way, extended to the worldly objects. The sense in which the world is 'incorporated' into practices should be initially differentiated from the way in which sounds and marks merely *convey* intentionality. This text conveys the intentionality of my assertions to you, but in no way do the pixels (or the ink of the printer) exercise authority over the correctness of what I say. Brandom's idea of 'lumpy concepts' seeks to catch a more robust sense in which the world partakes in discursive practices, somewhat like bats and balls 'partake' in baseball, where their purely material aspects, while in a sense contingent, are not *as* arbitrary as those of the signs we use in making assertions (MIE, 632, fn.). (Though, recall from the discussion at the end of 4.3 that the baseball analogy does not capture what is essential for Brandom's idea of incorporation, namely that the object itself can exert authority over the claims that are about it.)

How this works in practice can be appreciated by recourse to Brandom's algorithmic account discussed in Section 9.4. To illustrate, I shall here rely on Brandom's litmus paper example (BSD, 185). Consider the following causal chain of events:

1. The subject has a discursive attitude describable as a disposition to draw the inference 'If a substance tastes sour, it will turn litmus paper red'.
2. The subject has a perceptual experience of a substance that tastes sour.
3. The subject has a consequent perceptual experience of the substance turning litmus paper blue.
4. The subject loses her attitude-disposition to infer 'If some substance tastes sour, it will turn litmus paper red'.

In this example, we can see the *causal* entanglement of practices and the world. On the side of the practices, we have events (or states) (1.) and (4.), and on the side of the world, we have events (2.) and (3.). (Alternatively, we could replace, in this instance, the term 'practices' with that of 'abilities', for although in MIE, Brandom's official stance is that the relevant dispositions can only emerge in the context of intersubjective practices, he has later become less committal about this point; see BSD, 9, fn.) Of course, the whole chain of events is part of the same world, i.e. the distinction between discursive practices/abilities and the world is drawn from within the world when viewed in purely causal terms. A similar story on the side of action could be told where the subject's attitudes are the cause of changes in the world rather than themselves causally changed by how the world is (BSD, 178; MIE, 332-333).

The chain of events (1.) – (4.) above gives us a rudimentary grasp of how the world *causally* constrains the practices/abilities: paradigmatically by affecting our dispositions to draw inferences. Of course, not all such causal effects should be counted as having anything to do with how the facts of the world *justify* moves in discursive practices. How is it, then, that the world *normatively* constrains our practices/abilities? The key idea here is Brandom's commitment to *conceptual realism*, encapsulated by the notion inherited from Frege that facts just are true claims (i.e. what is *claimed* and not the *claiming* of it) (MIE, 327). Seen from the subject's own perspective, the claim to which she acknowledges commitment at (1.) turns out to be false in the transition from (2.) to (3.), i.e. in the face of the perceived fact that

there is a sour-tasting substance that turns the litmus paper blue instead of red. Here, the crucial difference between a claim merely *taken* as true and a claim that *is* true is made from within the practices/abilities as opposed to within the world: it is the difference between the subject attributing commitments (either to others or to her past self) and undertaking them herself (in the present). Since the subject-relative normative status of a claim as a fact depends on whether it is *only* attributed or *also* undertaken, and since the attitudes are already something involved in the causal realm of facts, the mechanism by which facts come to exercise authority over attitudes is given by the scorekeeping apparatus considered as causally integrated with the world in complex algorithmic ways.

The thesis under which Brandom pursues in this account is called 'conceptual realism'. The key claim here is that both facts and attitudes are conceptually structured, according to two different readings of the generic material incompatibility relation. On the side of the world, the concept of an object can be understood as 'repelling' incompatible properties under an *alethic* sense of necessity. On the side of the practices, subjects can be understood as 'repelling' incompatible commitments under a *deontic* sense of necessity. In Brandom's words:

It is *impossible* for one and the same *object* to have incompatible *properties* at the same time. But it is merely *impermissible* for one and the same *subject* to have incompatible *commitments* at the same time. (BSD, 191)

We can now better appreciate in what sense the world becomes 'incorporated' in or 'mediated' by discursive practices, following the litmus paper example above. Brandom's idea is that the succession of events (1.) – (4.) can be understood from two different *modal* perspectives, depending on whether it is described objectively as what happens or subjectively as what ought or may happen. The world and the practices are ontologically speaking two sides of the same event or process, structured in the generic modal sense of material incompatibility relation, which Brandom takes to be the key conceptual notion.

However, at this point, it would be equally correct to say that the practices are incorporated in or mediated by the world rather than the other way around. To make an already impressive amalgamation of theses more complicated, Brandom also pursues an explanatory order he attributes to Hegel, according to which the objective

side of alethic modal incompatibility *relations* must be understood and explained in terms of the subjective side of deontic modal incompatibility *processes* (TMD, Ch.6). However, as said before, in this work, I will not actively engage with this side of Brandom's project. My main claim is that conceptual realism is Brandom's key to how the objectivity of conceptual contents is *supposed* to be achieved.

11.6 Why Conceptual Realism Is Essential for Proper Objectivity

For one, Andrea Clausen (2008) argues that conceptual realism is non-essential and in fact a distraction for Brandom's aim of accounting for objective contents in terms of normative, discursive practices. The basic reason she considers conceptual realism (or 'the Hegelian piece') redundant is that she thinks Brandom's scorekeeping account alone can afford to explain proper objectivity and representational purport. The problem with that is that she does not adequately distinguish between attitude-transcendence and proper objectivity, i.e. between the negative claim that everyone could be incorrect in (some) of their assertions and inferences, and the positive claim that it is the world which determines the semantic incorrectness of assertions and inferences. Again, the bottom reason attitude-transcendence does not amount to proper objectivity is that, even if every subject in the practice necessarily *presumes* a difference between what is taken to be correct and what is correct, and that there is only one correct set of assertions and inferences everyone should acknowledge, it does not follow that it is the *world* that determines the identity of the set, or even that there *is* such a set. However, in detail, the fact is that *taking* the world to constrain assertions and inferences does not amount to the claim it *does*. Here's a telling excerpt of this *non sequitur*:

What we claim to be correct can always turn out to be incorrect. Put alternatively, this means that we rub ourselves against a resistant reality. Second, what is correct is supposed to be independent of what anybody or all take to be correct. Put alternatively, this means that we refer to one and the same world. (Clausen 2008, 217)

In fact, the reason our claims can always turn out to be incorrect, as far as the scorekeeping practice is concerned, may be that the contents are actually indeterminate or then determined by a Cartesian demon. And even if everyone agrees that what is correct is independent of what everyone takes to be correct, it remains possible that there is no reference to one and the same world.

Ronald Loeffler (2017) sees the problem of deriving (PO) from (AT) without further argument more clearly. Returning to the litmus paper parable, what Brandom wants to say is that *by* treating two of her commitments as materially incompatible with each other, the subject *takes* her commitments to be purporting to represent a singular object, namely the property of acid, for objects are (in part) defined as those entities, which repel incompatible properties in the alethic modal sense. Loeffler raises the question, however, of why we should interpret the subject as purporting to represent an *object* by taking two of her commitments to be incompatible, for on the face of it, we might equally well interpret her taking the incompatibility to amount to nothing more than a prohibition against endorsing two given assertion types (Loeffler 2017, 147). In other words, how does the intra-practice matter of which assertions are taken to be incompatible translate into the extra-practice matter of representational purport?

Loeffler's answer on behalf of Brandom is that although from our point of view as external theorists, the subject of the acid parable is not yet definitely purporting to represent anything beyond her practices or abilities; from the subject's *own* perspective, it appears that the acid itself serves as the external standard of her commitments, which hence purport to represent how things really stand with acidic substances (2017, 148).⁵⁰

The distinction between the native subject's own perspective and that of the external theorist's cannot, however, offer a sufficient reason to claim that the scorekeeping practice includes norms with representational purport or (PO) objectivity grading. The reason is, again, that the predicates 'takes to purport to represent an object' and 'purports to represent an object', or alternatively 'takes to be correct' and 'is correct', have distinct extensions, and claiming one does not entail the other. In particular, since Brandom's final major statement in MIE is that we are, in fact, engaged in the scorekeeping practice ourselves (the move he calls 'the collapse of perspectives'), any difference to the extent, which so starkly distinguishes between the native

50 Note that saying this is compatible with Brandom's insistence that although the subject is, from her own point of view, purporting to represent objects, the purport may be completely implicit in her practices or abilities in that she may not be able to explicitly assert that her commitments represent something external (Loeffler 2017, 149). The distinction between an implicit ability to do something that is independent of the explicit ability to *say* what one is doing is important to Brandom's pragmatist account of intentionality, though it is also largely orthogonal to the issues I am addressing here.

scorekeeper and her external interpreter, cannot hope to be adequate as an account of actual representational purport – if by 'actual', we mean whatever we do in purporting to represent objects. Applying Loeffler's response to our own case, even if it (necessarily) *appears* to us that we are responsive to the objects of the world when encountering incompatible commitments, it does not follow that we *really* are purporting to represent such objects.

Loeffler is sensitive to the tension internal to Brandom's view on semantic objectivity as it has developed over the decades. In particular, he notes that his later works appear to contradict the merely structural nature of semantic objectivity (which only really amounts to attitude-transcendence), owing to his historical extension of the scorekeeping practice. I also agree with Loeffler that modal realism, which, for Brandom, is near-synonymous with conceptual realism, is what provides the crucial element required to reach proper objectivity, which likely means compromising some of the key doctrines of MIE (Loeffler 2017, 177, 179). Briefly, the problem is that conceptual/modal realism as a metaphysical thesis is not entailed by the scorekeeping practice, yet in MIE, Brandom does nothing to give independent motivations for it. This means that the account cannot be 'expressively complete': it cannot be used to *say* what *we* are *doing* in terms of the scorekeeping practice alone (MIE, 641).

To end this section, I wish to reject one further argument, which seeks to establish representational purport in the scorekeeping practice without resorting to conceptual realism as an independent metaphysical theory. Bernd Prien (2010) argues that what is needed to ensure proper objectivity is a special norm called the 'principle of rational rectification' (PRR). The principle of rational rectification, which Brandom introduces in *Between Saying and Doing*, states that subjects are obliged to rectify the incompatible commitment they have committed themselves to. Indeed, as we already saw in this section, the principle in part defines the concept of a discursive subject for Brandom (BSD, 193). Prien claims that:

[p]ractices that include such a norm of rational rectification warrant an interpretation according to which the conceptual norms and thus the deontic statuses of the speakers are not determined by the deontic attitudes present in a community, but rather by the way the world is. Whenever a speaker runs into incompatible commitments because of the way the world is (for example, because there are sour-tasting liquids that do not turn litmus paper red), she is obliged to modify some of

the inferential relations she acknowledges. In order to make sense of this obligation, we have to assume that it is the world that determines what follows from what, and not the individual subjects, the experts, or the community as a whole. For even inferential relations accepted by the community as a whole have to be modified if this is the best way to remove an incompatibility. (2010, 455)

Prien claims that the PRR is a sufficient condition to warrant the properly objective status to conceptual contents in the discursive practices, for it is the only way to make sense of this obligation. It is difficult to see how that follows, however, for it is perfectly intelligible that everyone in the practice is obligated to rectify their incompatible commitments and that the world does not determine what commitments really are incompatible. Furthermore, it is not clear how precisely the world is supposed to *oblige* subjects to rectify their incompatible commitments other than in the metaphorical, causal sense of obligation (cf. Brandom 2008).

A similar point applies to another special norm, also mentioned by Prien, which we might call the *intersubjective* principle of rational rectification as opposed to *intrasubjective* PRR. The intersubjective PRR, first proposed by Loeffler (2005), states that different subjects A and B are obligated to rectify their commitments that are incompatible with some commitments of the other. For one, the intersubjective PRR seems to complicate Brandom's claim that we can define subjects as units of accountability *qua* subjects to intrasubjective PRR. If PRR is extended from intra- to intersubjective incompatibility relations, are we to conclude that two distinct subjects can form a singular unit of discursive accountability?

More acutely, it remains unclear how PRR in either its intra- or intersubjective versions is supposed to entail that subjects really are responsible to the world in what concerns the correctness of their commitments. For the issue of *in virtue of what* commitments really are incompatible is independent of whether, and in what sense subjects are obligated to rectify their incompatible commitments. Even if it is the world that somehow non-metaphorically obliges the subjects to rectify their incompatible commitments, something which Brandom explicitly denies (2008), it is a different matter to establish whether the world also determines (and does not merely appear to determine) which commitments are incompatible. So, PRR alone does not entail that the scorekeeping practice that includes it also includes norms with representational purport.

11.7 Summary of the Section

The key idea that this section pursued is that there is a mismatch in Brandom's characterisations of the criteria in terms of which the scorekeeping account should be judged as adequate for what comes to semantic objectivity. On the one hand, Brandom relies on the structural notion of semantic objectivity, which amounts to the level of attitude-transcendence (see Section 11.1). Effectively, every scorekeeper distinguishes between what is correct and what is merely taken as correct, even in their own case. On the other hand, Brandom's more informal characterisations of semantic objectivity include the notion that it is the world of objects somehow incorporated in the practices that determines, at least partially, the correctness of the claims that are about it. This is the level of proper objectivity, which, I argued above, is distinct from attitude-transcendence because (AT)-norms are compatible with the possibility that it is not the world of objects that determines what is really correct and what really follows from what.⁵¹

My proposed solution to this mismatch was that, already in MIE, Brandom's best bet to account for proper objectivity is to explicitly embrace conceptual or modal realism as an independent metaphysical thesis, which dovetails well with the scorekeeping account. However, in this section, I did not go into the reason this is Brandom's best bet, or why he, in fact, appears to rely on it in MIE. An obvious problem here is that the truth of conceptual realism is independent from the scorekeeping account, thus requiring an independent metaphysical justification, which MIE does not even touch upon. The question arises of whether Brandom's broader aims would be better served by simply giving up the notion of representation completely, thus settling for the level of attitude-transcendent objectivity.

The reason Brandom does not want to give up the notion of representation and the level of proper objectivity in MIE, I argue below in Section 12.5, is that this is his strategy for arguing for the truth of normative phenomenalism. Ultimately, Brandom wants to hold onto his synthesis between phenomenalism about norms and normative phenomenalism, and for that, he needs the level of proper objectivity and a more robust notion of semantic objectivity than what attitude-transcendence can

⁵¹ Much of the argumentation in this section is parallel to my work (2021a).

yield. As we shall see, this is a key part of his more fundamental claim that the norms implicit in practices are *genuine* norms. However, below, I shall also argue that this strategy does not work as Brandom intends.

12 PRAGMATIC OBJECTIVITY

12.1 On Bindingness

In Section 11.1, I described the standards of semantic objectivity that Brandom sets for himself to fulfil in MIE. The two broadest characterising traits of the three-tiered objectivity grading are, first, *phenomenalism about norms* and, second, a negative view of what it means for conceptual contents to be *determined*. So, for descriptive empirical conceptual contents, which arguably is the domain we are most interested in as semantic theorists, proper semantic objectivity is reached when the content of the norms is determined by the world – as opposed to by the community or by some Cartesian demon – in the sense that the whole community could be judged to be incorrect in their grasp of the inferentially articulated contents of the conceptual norm.

A key question about semantic objectivity for Brandom, then, concerns the *source* of the authority, which is to determine the correctness of adopting discursive attitudes. Although the methodological commitment to phenomenalism about norms states that the ultimate normative authority is to be sought only from discursive attitudes, that is compatible with the authority being in a way extended to certain worldly objects. The objects come to be incorporated in, or mediated by, the discursive practice and normative attitudes, which introduces a notion of representational purport into the practice. How this mediation is supposed to work was explained in Section 11.5. Presently, I focus on introducing the pragmatic side of norms' objectivity, by which I shall mean their *force* as opposed to their contents.

In contrast to semantic standards, the evaluative criteria fit for pragmatic objectivity are harder to pin down in Brandom's writings. Here, the central topic of authority, the domain of its reign, are not descriptions but *prescriptions*, and in particular, the *normative force* such prescriptions exercise on discursive subjects. By 'prescriptions', I mean to capture both obligatory and permissive forms of normative force usually

denoted by the terms 'ought' and 'may' respectively, including proscriptive force. The key form that normative force takes, for Brandom, is *responsibility* or the ought of obligation, for as Frieder Vogelmann (2020, 929-930) observes in his interesting discussion, although the authority-articulating concept of entitlement is *semantically* on equal footing with the responsibility-articulating concept of commitment, *pragmatically* responsibility is used to define authority. For to be authorised to authorise others to assert a claim is to be treated as subject to the conditional task-responsibility to justify the commitment if appropriately challenged to.

The sense in which prescriptive attitudes and claims can be said to be 'pragmatically objective' will be elaborated in the next subsection, to which the present one serves to pave the way. To anticipate, I believe we can distinguish three different models in Brandom's works, according to which discursive normative force becomes *binding* on subjects: the Kantian autonomy model, the Queen's Shilling model, and the Hegelian mutual recognition model. Roughly, 'bindingness' (or Kant's *Verbindlichkeit*, sometimes translated as 'validity' (Brandom TMD, 21)) means the conditions in which prescriptive attitudes or claims can be said to be correctly adopted or made. The term 'bindingness' has such massive import to pragmatic objectivity that it is worthwhile to devote this first subsection to its various meanings in Brandom's works.

The first clarifying distinction relating to bindingness is between the predicate expressions 'being subject to a norm' and 'being sensitive to a norm' (Brandom TMD, 12). Briefly, to be subject to a norm means that one's behaviour is correctly assessable by it. For example, I may assess a director's work on a film, according to the aesthetic standards of the genre to which it belongs. But this is not yet to say that the director is, in any way, *sensitive* to my assessment or the standard behind it. 'Sensitivity' should be understood, I believe, in the causal sense in which the director is or would be *motivated* to behave by my assessment and the standard behind it. The crucial point here is that being sensitive to a norm is not any kind of a condition for being subject to it: one may be subject to a norm, e.g. to a law, even while one is not in the least motivated by it or is even motivated to behave against it. While Brandom does concede that in the discursive case, there might be a minimal necessary amount of sensitivity required on the subject's part for them to be properly assessable by discursive norms, the issues of sensitivity and assessment should generally be kept

apart (RP, 3). This means that the primary sense of bindingness for Brandom is assessability according to the norm.⁵²

Another important point about bindingness is the distinction between a norm *being* binding for a subject and being *taken* to be binding; in other words, a norm being *correctly* taken as binding and being taken as binding. As we shall later see, much as concerns pragmatic objectivity of normative force will hinge on the extent to which acts of bindingness can be performed *incorrectly*.

A third important point about bindingness relates to a kind of *qua*-question: *qua* what is a subject bound by a norm? At least two answers are available. First, a discursive commitment to *p* can be binding on a subject *qua* her role as an asserter, either because an assertion of hers immediately expresses a commitment to *p* or then because *p* is undertaken as an inferential-committive consequence of another assertion. Second, a norm can become binding on a subject *qua* a social role the subject occupies. For example, a subject occupying the role of a doctor is subject to the Hippocratic oath to care for her patients. We should also distinguish between practical and theoretical commitments, i.e. being committed to act and being committed to a claim; a difference that corresponds to the roles of an actor and a believer (MIE, 245). More relevantly in the case of specifically discursive norms, Brandom wants to argue that certain norms, so-called *fundamental discursive norms*, are binding on subjects *qua* their roles as scorekeepers (Brandom 2008). Among these fundamental discursive norms (FDN) is the intrasubjective PRR we encountered in Section 11.6, as well as the conditional task-responsibility to justify commitment to a claim if appropriately challenged to.

⁵² Alternatively, we could understand 'sensitivity' not only as motivational force but as a kind of specific qualitative feeling or experience. This is what Ginsborg does when she claims that what distinguishes the dispositional responses of a parrot and a human is a special sort of understanding on the human's part (see Section 6.1). Curiously, Brandom also asserts that what differentiates the parrot from the human 'is a kind of *understanding*,' although he makes it clear that by this, he means a specific practical capacity and not an inner qualia (MIE, 88-89). So, sensitivity to norms covers not only the *motivation* but also the (graded) *ability* to follow their dictates.

12.2 Three Models of Discursive Sovereignty

In the previous subsection, I distinguished between certain elementary meanings of the term 'bindingness' in Brandom's works. In this one, I focus on the question *in virtue of what* any discursive norms become binding on subjects. As already noted, Brandom endorses the general methodological commitment to phenomenalism about norms, according to which the origins of discursive norms should be sought only from the attitudes of subjects. The three models that I will consider – Kantian autonomy, Queen's Shilling, and Hegelian mutual recognition – all submit to this broad principle, differing only as to *whose* attitudes should count most in determining when and how a norm becomes binding, and to whom. For this reason, all three models can be understood as more fine-grained variants of phenomenalism about norms. As a further terminological note, I shall use the term 'sovereignty' to mark the power to make (discursive) normative force binding, and 'genuine' to mark the success of acts of bindingness, where 'success' means being correctly taken as binding.

To begin with, the Kantian autonomy model can be summarised as the claim that only the subject herself has the necessary and sufficient authority to bind herself by genuine normative force. The defining idea of autonomy, which Brandom takes Kant inherited from Rousseau, is that only the subject's own, self-conscious attitude of *acknowledging* a commitment suffices to make it binding (MIE, 52; TMD, 219). However, this sense of bindingness is compatible with the subject's ignorance as to *what* precisely she has committed herself by acknowledging a norm, i.e. the content of the obligation that she has chosen to fulfil. In corrigibility in the pragmatic affair of acknowledging a norm does not translate to incorrigibility in the semantic affair of determining what content the norm has beyond the acknowledgement of the subject herself, so that correct assessments of the subject's *success* in fulfilling her acknowledged obligations do not have to coincide with the subject's *own* assessment for the bindingness to be genuine.⁵³

⁵³ It probably goes without saying that the 'Kantian model' ought to be decapitalised, for it is so slim in content as to make mockery of Kant's actual thoughts on the matter. However, as an expositional device, the model serves its purpose.

The opposite of Kantian autonomy can be aptly called the 'Queen's Shilling' model. The name refers to the early modern British navy practice of recruiting illiterate citizens by offering them a royally minted coin, the acceptance of which by law authorised the recruiting officers to hold the subject responsible for joining the navy. Here, the source of discursive sovereignty lies with the attitudes of others in *attributing* the commitment to the subject of the norm, irrespective of the subject's own knowledge and volition about the matter. Whereas the Kantian autonomy model defines the genuineness of any assessment of bindingness (albeit not the correctness of content) in relation to the subject's acknowledgement, the Queen's Shilling model defines acknowledgement itself as an act by the subject, which authorises her audience to attribute the commitment to her, where authorisation is understood as whatever is the actual, working practice in the community (e.g. accepting a royally minted coin from recruitment officers). As with the autonomy model, the Queen's Shilling model also affords the crucial distinction between attributing a commitment and its content.

In MIE, Brandom explicitly opts for the Queen's Shilling model as the paradigm for how normative force becomes binding on subjects, which shows in that he takes the attribution of commitment to be the fundamental explanatory attitude as explained in Section 9.2. Basic refinements to the Queen's Shilling model concern primarily the sanctions (internal vs. external) and the kinds of normative statuses (not only commitments but also entitlements) available to the practice (MIE, 159, 164). The decision to privilege the second (or in fact, the third) person perspective is also clear in MIE's explicit aim to account for the interpretive stance of attributing original intentionality to a community, by which Brandom means interpreting the community as interpreting each other as attributing discursive commitments and entitlements to each other independently of an external interpreter (MIE, 61). We shall return to original intentionality in Section 12.6 in the context of how Brandom can handle animal intentionality.

To further elaborate them, I shall now briefly discuss the respective problems of the Kantian autonomy and Queen's Shilling models, and by that token, shed some light to their Hegelian synthesis. In this section, I cannot elaborate on how the Hegelian mutual recognition model seeks to remedy the faults of the two other models other than in the most abbreviated of terms: the topic will be picked up in Chapter 3. The

principled way in which the Hegelian model of mutual recognition seeks to overcome the incompatibility between the two other models is to regard both the attitude of acknowledgement and that of attribution as equally important in making the discursive normative force binding (ST, 278, fn.).⁵⁴

The key complaint about the Kantian model, which Brandom himself considers is that it simply posits a metaphysical power for discursive, rational subjects to bind themselves by norms without even attempting to explain how this power could have originated or what in the actual physical world might realise it (TMD, 13; RP, 81). The more particular Hegelian aspect of this complaint is not just that the power to invoke norms would be 'spooky' as considered from a purely naturalistic perspective, but that it is a power without a history or a constitutive social dimension; a power that humans recognisably have, but of which they have no idea how it has come about, what its nature is, or whether and how it might one day cease from existing.

While the Kantian autonomy model struggles to explain in phenomenalist terms how discursive sovereignty has originated, the Queen's Shilling model faces the opposite allegation of collapsing norm-governed practices into *de facto* power relations and contingently existing social practices. Of course, just as the complaint about the metaphysics of sovereignty can only beg the question against the ardent Kantian, the Queen's Shilling model will appear nothing short of commonsensical to a Hobbesian or Machiavellian (and in contemporary times, Foucauldian or Bourdieuan) thinker, for whom our faculty of reason is nothing more or less than conditioned by the social, and in the final analysis political, reality (to exaggerate a bit). The particular problem for Brandom is that he is not in principle willing to endorse the social positivism inherent in the Queen's Shilling model, but wants to preserve the Kantian commitment to the autonomy of reason, not only as concerns its contents but also its force. Hence, the critical reading of MIE by Robert Pippin along these lines poses an internal dilemma for Brandom instead of an external attack (Pippin 2005).

⁵⁴ Moreover, it is not just the normative attitudes of attribution and acknowledgement that are treated as equals, but also the normative statuses of responsibility and authority. For in the Kantian model, autonomous authority is taken to be the primary status that explains the ability to become responsible, whereas in the Queen's Shilling model, this very authority is explained by the status of responsibility attributed by others as was mentioned above.

In what follows my purpose is to reconstruct Brandom's argument in MIE that strives to synthesise the commitment to the Kantian autonomy of reason with the evidently reductive Queen's Shilling model that privileges the attitude of attributing a commitment as sovereign. Briefly, in my view, the key moment in the synthesis is to argue for the pragmatic objectivity of force via the semantic objectivity of contents, which in turn are 'conferred' on expressions by the attitudes. In other words, while subjective, pragmatic, normative attitudes explain the emergence of objective semantic norms, the objective semantic norms in turn explain the pragmatic objective force that binds the attitudes. I shall further argue, however, that the reconstructed synthesis faces some grave problems.

12.3 The Problem of Reason's Autonomy

Before moving on to discuss how Brandom aims to resolve the incompatibility between the Kantian autonomy of reason and the reductive Queen's Shilling model in MIE, I find it useful to motivate the precise problem in more detail. The initial idea is that we can use the semantic objectivity grades, presented in Section 11.1, to discuss objectivity also in the pragmatic dimension, at least in the cases of the Kantian autonomy and the Queen's Shilling models, which are the focus of these subsections. Indeed, I take it that the change of one word suffices to transmute the objectivity grading into a pragmatic key:

(AI_{pr}) A norm n is attitude-immanent for community C iff it is not possible for everyone in C to be mistaken about the *force* of n .

(AT_{pr}) A norm n is attitude-transcendent for community C iff it is possible for everyone in C to be mistaken about the *force* of n .

The subscript (pr) will henceforth designate the specifically pragmatic type of objectivity grade as opposed to the semantic variant (sm) . In this terminology, then, discursive sovereignty comes in at least two different degrees of pragmatic objectivity. The question that follows is which type the Kantian autonomy and Queen's Shilling models respectively involve, and what problems this may engender.

As it will turn out, this rendering of pragmatic objectivity grades will require specifications later in Section 12.7, but it will do for now.

The pragmatic objectivity grade available to the Kantian autonomy model is readily appreciated to be attitude-immanent, for not everyone can be intelligibly taken to be in error about whether a given commitment is in force for a subject. Namely, the subject herself cannot be in error simply because it is her self-conscious attitude of acknowledging commitments as binding, which immediately *makes* them binding. Likewise, when we turn our attention to the Queen's Shilling model, it is equally obvious that the whole community cannot be in error in attributing a commitment to a particular subject based on some performance of theirs simply because it is the standing, working practice that makes the attributions authoritative (see especially Pippin 2005, 395).⁵⁵

However, although both the Kantian autonomy and the Queen's Shilling models fall in the purview of attitude-immanent norms in terms of the pragmatic objectivity grading of the normative force, the latter kind is more objective (or at least more nuanced) in the sense that no *singular* attitude is sovereign but only the collectivity is, however that is understood (e.g. as unanimity or as the common judgement of some select group of experts within the community). In contrast, in the Kantian model, sovereignty is always singular. What conjoins these views beyond their differences to the singularity/plurality of the instance issuing authority is the *indivisibility* of sovereignty as such that is implied. Indivisibility in this sense does not hinge on the *number* of subjects required to make a judgement about normative force as binding, but on the act of judging itself. Whether it is considered as issued by a singular or a plural subject, as an acknowledgement or as an attribution⁵⁶, the act that makes force

⁵⁵ Fossen (2014, 389-390) objects to Pippin's reading by pointing out that subjects may pretend to take the attributions of commitments seriously, which he takes to entail that subjects may intelligibly take a whole practice to be corrigible about what counts as properly acknowledging a commitment. However, as I already pointed out in Section 11.2, the issue of whether the actual questions of sovereignty are contestable or not is unimportant to their objectivity grading_{sm}: e.g. marriage norms can be both contested and (AI_{sm}) at the same time. Similarly, even if everyone merely pretends some practice to be genuine, it does not follow that it *would not be genuine if everyone stopped pretending*; hence that it is only (AI_{pr}) by its objectivity grading_{pr}.

⁵⁶ The dimensions of singular/plural, as well as the dimensions of acknowledgement/attribution, can be cross combined without this affecting the indivisible quality of sovereignty. A non-discursive example of a singular instance capable of sovereignly attributing commitments to others would be the

binding is considered by both accounts to be incorrigible, capable of instituting normative force *immediately*, i.e. sufficiently and necessarily by itself.⁵⁷

Again, there is more to be said than here can be about the nature of pragmatic, normative force as depicted by these models. For one, as with semantic objectivity, the incorrigibility of the whole community does not entail that the community cannot suffer from internal strife when it comes to disputing claims about discursive authority, e.g. when a performance rightly counts as an assertion, whether one should be interpreted as having *requested* or *demanding* for something, etc.⁵⁸ One crucial thing to be noted here is that simply because sovereignty is deemed attitude-immanent by its pragmatic objectivity grade by these models, it does not follow that it is in the same *ontological* category with social norms. For it is compatible with the Kantian model that (discursive) sovereignty is *sui generis* metaphysically, a transcendental power of the rational subject (in the sense of 'a condition of possibility') and yet attitude-immanent_{pr} by its pragmatic objectivity grade. On the other hand, neither does it follow that discursive sovereignty would be incompatible with the ontological or metaphysical status of merely social normativity – the autonomy of reason has to be won by the toil of arguments.⁵⁹

But what *is* 'reason's autonomy', and what problems inhere in the concept? The problem of reason's autonomy as it presents itself to Brandom is, briefly, that he evidently does not want to consign discursive sovereignty to the same basket with social norms, such as traffic regulations or tort law, although in view of the Queen's Shilling model, this might seem like the most natural ontological zip code. At the same time, he accepts the Hegelian complaint that discursive sovereignty without a history and roots in social practices is an idle label, amounting to a form of

Pope, while an instance of a plurality capable of sovereignly acknowledging commitments would be a citizens' assembly declaring independence.

⁵⁷ In this terminology, then, one way to describe the Hegelian synthesis via the mutual recognition model is as an account of divisible or shared (discursive) sovereignty.

⁵⁸ Lance & Kukla (2009) are an abundant source for these sorts of questions and answers for them.

⁵⁹ Implicit in my discussion is that the autonomy of reason must be, in the primary sense, given an ontological or metaphysical reading, which is probably not an incontestable supposition – the primary sense of autonomy could well be ethical, for instance. What I do take to be self-evident though is that the Kantian autonomy of reason is incompatible with the ontological/metaphysical status as merely social normativity, if by that one means whatever may be empirically studied.

primitivism denying that the power could in principle be explained as a socio-historical artefact.

So, assuming that Brandom is not a (Kantian) primitivist about normative force, it must be possible to elaborate on *why* it is that certain takings and treatings by discursive subjects succeed as genuine attributions of commitments and entitlements, and why others possibly fail to do so. The possibility of an attribution of a commitment being erroneous regarding the *content* attributed does not suffice here, for even if my attribution of (say) an incompatibility between your commitments p and q is semantically incorrect, in the sense that p and q are not really incompatible, it does not follow that my taking was not a *genuine* attribution of a commitment to begin with. If it was not, then I could not be responsible for my incorrect attribution, which I clearly am, assuming that it really is incorrect.

However, why would it not be possible to understand attributions and acknowledgements as incorrigible (at least in some ideal conditions) and still *genuine* normatively speaking? This is, after all, the founding premise of the Kantian autonomy model of sovereignty: the subject herself generally cannot intelligibly be taken to fail in her acknowledgements of norms, which does not contradict the genuine status of her sovereignty: her acknowledgements, simply as a rule, succeed in their purpose. Indeed, there is no formal contradiction to this combination of commitments. But once we take on the further task of explaining how the sovereign power to bind oneself with norms has originated in the course of natural and human history, it becomes difficult to justify an account of an act, which cannot generally *fail* to succeed (or be genuine) simply because it appears to be an essential element in our ordinary concept of action that it can generally (and all too often) fail. This reveals a central facet in which the Kantian insistence that the sovereign acknowledgement of a commitment is a *special* act amounts to primitivism about discursive normative force: it is made special by inherent immunity from failure.

The overall lesson here comes across as a criterion of adequacy for any account that aims to explain discursive sovereignty as a socio-historical achievement, namely that it must be in principle possible to explain how the acts of exercising sovereignty can fail, not just in *outré* and *recherche* examples, but as a matter of everyday happenstance. This condition dovetails nicely with the negative account of

objectivity to which Brandom submits on the semantic side, where the measure of a norm's objectivity_{sm} is the degree to which assessments may be found incorrect according to it, with the most objective norms being those whose contents everyone could get wrong. Likewise, the account of genuine discursive sovereignty is complete not simply when we can explain the conditions in which the norms get a grip on subjects, but also when we can identify the conditions in which they come to lose that grip. The point is duly appreciated by Pippin (2005, 385).

In order to make this criterion of adequacy more exact, I propose a 'genuineness test' for normative force on the model of the semantic objectivity grading discussed above. Simply put, for a norm to pass the genuineness test, it must be shown that its pragmatic objectivity grade is (AT_{pr}) as opposed to (AI_{pr}). It should be observed that as the test captures only a necessary and not a sufficient condition for genuine discursive sovereignty. For it remains compatible with a norm being (AT_{pr}) that nothing determines the genuineness of normative force, even though everyone takes something to determine it; a possibility that effectively amounts to error theory (or rather, 'failure theory' since we are talking about the pragmatic dimension). A sufficient condition for genuine discursive sovereignty would in turn amount to an equivalent of the (PO_{sm}) grading of a norm on the pragmatic dimension. Whatever account would fulfil *that* condition is not discussed here, but only in Chapter III. Here, I shall only focus on the necessary condition (AT_{pr}) and whether Brandom's theory in MIE can account for it. My argument below in Section 12.5 will be that it cannot.⁶⁰

Why is it though that passing the genuineness test requires a norm to be (AT_{pr}) as opposed to (AI_{pr}) if what is sought are only conditions in which *some* attributions of commitment may fail to be genuine? After all, both the Kantian autonomy and Queen's Shilling models allow for some attributions to fail. For the Kantian, every attribution of a commitment to me) that contradicts my own acknowledgement of the commitment (if not my acknowledgement of the commitment's content) fails to be genuinely normative because it does not respect my metaphysically objective

⁶⁰ It is possible, at least in the abstract, that pragmatic and semantic objectivity are asymmetrical in the sense that while (AT_{sm}) is insufficient to account for determinate *content*, (AT_{pr}) is sufficient to account for determinate *force*. This would imply a situation where it is determined when an act fails to be sovereign without it being determined what would count as success of sovereignty. This counterintuitive yet interesting possibility will be discussed in Chapter 3.

normative status as an autonomous, rational subject. Although I may be actually treated by everyone as someone committed to some claiming, I may never be *correctly* so treated without my acknowledgment of the claiming. Likewise, the officers in the Queen's Shilling model may fail to genuinely attribute the commitment to join the navy to someone if the coin they offer was not actually manufactured in the royal mint, making it counterfeit. In other words, both models are perfectly capable of accounting for *local* failings of attributing genuine commitments – why is the more demanding *global* requirement for failure at the level of the community needed?

There are two reasons, one for each model. In the Kantian case, the fault lies with its primitivism, which rejects all attempts to give socio-historical explanations for why individual subjects are sovereign in the discursive sense, making the account unable to explain why certain performances do count as genuine attributions of commitments while others do not. Supposing that Brandom is interested in explaining that he cannot be satisfied with the merely local failings of attribution.

The contrasting reason local failings will not do with the Queen's Shilling model is that the whole community cannot be wrong regarding its standing practices that *define* what counts as acknowledging commitments, although, of course, potentially everyone can be mistaken (at least sometimes) whether the defining criteria are fulfilled in context or not. Assuming that the coin accepted is genuine (along with other conditions decreed by the law and customs), hence must be the commitment, and it changes nothing to admit that some counterfeits may be so good as to pass everyone's vigilance. The model is inherently reductive regarding normative force precisely because it identifies the community's *actual* practice, which can be fully described in dispositional and causal terms, as capable of instituting genuine discursive sovereignty. But that just means giving up the commitment to the Kantian autonomy of reason, which understands genuine normative force to be incompatible with merely social normativity exhibited by the brute natural fact (if there is such a thing) of community agreement.

To summarise, the reason Brandom is committed to the more demanding form of the genuineness test in the global (AT_{pr}) sense is that he wants to hold onto phenomenalism about norms and normative phenomenalism on both semantic and pragmatic dimensions. The incompatibility of the Kantian model with global failings

of sovereignty amounts to giving up (a central aspect of) phenomenalism about norms, for the force of some norms is considered primitive in socio-historical respects. Meanwhile, the incompatibility of the Queen's Shilling model with global failings of sovereignty amounts to giving up (a central aspect of) normative phenomenalism, for the force of norms is ultimately reduced to the natural fact of community agreement. If Brandom wants to combine the two models, which I believe he does, his account needs to pass the stronger version of the genuineness test.

12.4 Original Error

The considerations above cue us to see the more general fault (from the synthesising perspective) that conjoins the otherwise opposite Kantian autonomy and Queen's Shilling models. In a slogan form, both models take *pragmatic incorrigibility* to be a condition of possibility for *semantic corrigibility*. That is to say, in order for there to be semantic errors, in the sense of the potential for representational success, according to these models, we must posit some region in the pragmatic domain (i.e. in the domain of *takings* and *claimings*) where error is not possible for everyone. For the Kantian autonomy, the region is found in the metaphysical, transcendental power of subjects to bind themselves with norms, considered primitive in relation to socio-historical reality. For the Queen's Shilling model, which gives primacy to socio-historical practices, the incorrigibility is in the non-discursive layer of social norms, such as who counts as properly married or what punishment ensues from entering the sacred hut without a holy leaf. This 'ur-normativity' forms the primitive base of the scorekeeping practice, and it is (AI) in both objectivity grading statuses, so not everyone can be in error about the content of these norms or whether they are in force for the community members.

It is now curious to see that this general architecture, where primitive incorrigibility enables explanations of more refined corrigibility, is also what unites Descartes and Kant – the very antonymic figures in Brandom's history of philosophy. In effect, what Descartes attempted to do in the domain of epistemology, Kant repeats in the domain of normative pragmatics. For Descartes attempted to lay the groundwork for scientific knowledge, or more generally, our ability to veridically represent *res*

extensa, based on an internal, mental sphere of appearances or seemings where error is not possible, following the familiar fact (though not an uncontroversial one) that phenomenal predicates, such as 'seems red', are not Iterable. If something seems to seem red to me, it just seems red; hence, there is no such thing as 'not *really* seeming red', or *failing* to seem red.

Arguably, the most devastating objection to this epistemic strategy originates from semantic and, more accurately, pragmatic considerations. For as Wilfrid Sellars in Brandom's (TMD, Ch.12) influential reading of *Empiricism and Philosophy of Mind* argues, the semantic content of phenomenological predicates, such as 'seems red', are not *pragmatically autonomous* in relation to predicates, such as 'is red'. To cut a complicated argument short, no one can be in a position where they know how to apply the phenomenalist predicate 'seems red' without already knowing how to apply the extensional predicate 'is red'. The key reason is that applying the phenomenological predicate 'seems red' effectively has the pragmatic significance of *suspending* one's judgement whether the perceived object *really is* red or not, and that it is impossible to suspend a judgement unless one already can *make* judgements, i.e. apply the predicate 'is red'.

An alternative way to describe the central problem for the two models of discursive sovereignty that Brandom tries to synthesise in MIE is that neither can afford a notion of *original error* in the pragmatic sphere of normative force that would be independent of some primary incorrigible region – in Kant's case, the metaphysical power, and in the Queen's Shilling's case, the non-discursive norms. However, it boggles the mind to just wonder what exactly such 'original error' would entail. For what seems to be required is a notion of error such that it *precedes* the determination of correct applicability in the form of incorrigibility. The semantic case is comparable to the possibility of pragmatic asymmetry between the failure and success of discursive sovereignty mentioned above. For what original error effectively implies is analogical to *original sin* in the sense that one could be found genuinely guilty of crime/error without it being the case that it is possible for one *not* to be guilty, or to be correct in the application of a concept. Furthermore, the impossibility is not of the contingent kind to which Brandom apparently alludes to in BSD (191), where our 'original discursive sin' is the inability to compute all the possible incompatibilities and consequences of our undertaken commitments. Rather,

original error implies the possibility of an application of a term counting as determinately incorrect without the converse possibility of it counting as determinately correct; there is no logical *option* of being correct, only that of being not-incorrect. In short, we cannot redeem ourselves of original sin/error because of some contingent human failure, but because being subject to original sin/error just is what being a discursive subject means.⁶¹

I believe the shape of this problem is already greyly visible in Brandom's definition of material incompatibility relation, which states that two commitments are incompatible iff commitment to one precludes entitlement to the other and vice versa. As was noted in the early reception of MIE, Brandom nowhere explains what 'preclude' means (Rosenkranz 2001, 235). Is it to be understood in the causal-dispositional or normative sense? If the former is right, then presumably, we only arrive at a vague notion of what it means for scorekeepers to *take* two commitments to be incompatible. However, in the latter case where 'preclude' is read normatively as unjustifying the combined undertaking of two incompatible commitments, we are left wondering how the route from pragmatic abilities is supposed to entail *this* notion of justification.

I think Brandom is aware of the problems regarding the pragmatic objectivity of normative force in MIE that I have been tracing these past few sections. What, then, is his solution? As I argue in the next section, I believe his answer is to argue for the pragmatic objectivity of force by arguing for semantic objectivity of contents, for which he has argued earlier by the force of the said norms.

12.5 Why Semantic Objectivity Does Not Entail Pragmatic Objectivity

In Section 11.5 above, I characterised how Brandom's normativist, pragmatist strategy can explain the possibility of properly objective grading_{sm} for some

⁶¹ The idea of original error in both the pragmatic and semantic domains is, I believe, one of the crucial blank spots on the map that Brandom seeks to draw clear in ST. For he notes, already in TMD, that for Hegel, material negation seems to be the 'fundamental content-articulating notion' (TMD, 223, fn.29).

conceptual norms, or alternatively, how these norms can have contents with representational purport. As an account of semantic objectivity, the theory has a lot to say for itself, although it has also been vehemently criticised and is not free of problems. In particular, I think the major weakness of the account lies in giving independent motivations for the truth of conceptual realism, for Brandom's claim that the world can constrict discursive practices both causally and normatively hinges on the idea that the world itself is conceptually structured, thus able to mediate partake in the practices.

However, what concerns me here is the evaluation of the proposal on merits other than semantic, namely the pragmatic issue of genuine normative force. My argument now is that the account for semantic objectivity, even if successful, does not *entail* a corresponding account on the pragmatic side regarding the objectivity of normative force in the sense of passing the strong genuineness test described in Section 12.3.

Let me start by motivating the idea, which is not self-evident, that semantic objectivity is *supposed* to entail pragmatic objectivity to begin with. I already established that since MIE, Brandom's ambitious goal has been to combine the Kantian thesis about the autonomy of reason with a non-primitive reading of phenomenalism about norms, i.e. to understand the normative force of reason as both irreducible to non-normative powers, such as causality and dispositions, while also as something instituted by our social, historical discursive practices. This is not yet to say anything about how the trick is to be done, namely whether semantic objectivity plays any role in it.

To be sure, there is no literal smoking gun available in Brandom's writings that would commit him to the explanatory strategy I ascribe to him. But there are certain key excerpts in MIE, which I think at least strongly allude to this direction:

Interpreting a community as exhibiting original intentionality is taking its members to adopt the discursive scorekeeping stance towards each other. The content-conferring norms and proprieties that an interpreter who attributes discursive scorekeeping practices takes to be implicit in them have a number of important structural features. Central among them is the fact that the conceptual norms implicit in the practices attributed to a community outrun the nonnormatively specificable behavioral discriminations members of that community are disposed to make. For this reason, conceptual norms can be understood as *objective*, and so as binding alike on all

members of a discursive community, regardless of their particular attitudes. (MIE, 631)

The last sentence is crucial, for Brandom clearly appears to connect the (semantic) objectivity of conceptual norms with their 'bindingness alike on all members of a discursive community'. This is important because Brandom frequently uses 'bindingness' to mark the pragmatic affair of a norm's grasp on subjects, or the force it exerts on them (e.g. MIE, 5, 50-51). Perhaps even more revealingly, a few pages earlier, he writes that:

Talk of inferentially articulated contents is a way of talking about implicit norms governing deontic scorekeeping practice; this is the cash value of the claim that conceptual contents are *conferred* by such practice. But since commitments must be individuated at least as finely as their contents, if those contents are determined only by how it is *correct* to alter and acquire deontic attitudes, the commitments themselves must be understood as instituted also by proprieties of scorekeeping, rather than by actual scorekeeping. The scorekeeping account incorporates a phenomenalist approach to norms, but it is a *normative* phenomenism, explaining having a certain normative status in effect as being *properly* taken to have it. (MIE, 627, footnote omitted)

The main explanatory line of MIE is from the norms instituted by practices to the propositional, representational, conceptual contents conferred by the said normative practices. In contrast, Brandom remarks upon the reverse line where the conferred objective contents explain what it means for genuine commitments (i.e. normative force) to be instituted: not by how the score is actually kept, but by the proprieties of keeping it. In other words, the success of phenomenism about norms in explaining how semantic objectivity in the (PO) sense is possible entails the success of pragmatic objectivity regarding normative force. But that, I claim, is precisely the strategy for synthesising Kantian autonomy with the Queen's Shilling model that I ascribed to Brandom above.

Assume now that my exegetical claim is correct, and that Brandom seeks to argue for the institution of genuine normative force via the conferral of properly objective conceptual contents. What reason is there to doubt the success of this claim, assuming that proper objectivity is secured on the semantic side?

The main reason I have in mind has to do with the mediation/incorporation strategy discussed in Section 11.5. As we saw there, Brandom's argument for how objects of

the world can come to exert authority on our commitments depends, first, on the fact that objects of the world are causally integrated with our attitudes and experiences, considered as dispositional, algorithmic states and processes. The attitude-disposition to endorse inference from p to q is (usually) causally sensitive to the (perceived) fact of whether the fact p causally entails the fact q . Second, facts in turn are only intelligible, according to Brandom, based on the subjective practices or abilities of acknowledging and attributing commitments. The subject takes her attitude to be responsible to the facts, and thereby comes to give some worldly objects the authority over the actual contents of her attitudes.

However, although the worldly objects can be said to exert authority over *what* is correct, they do not *oblige* the subjects to acknowledge (only) correct commitments; not except in the metaphorical causal sense in which facts may oblige the revision of our attitudes about them. The mediation/incorporation model cannot account for genuine normative force simply because there is no such thing in the world to be incorporated or mediated by our attitudes, whereas there are conceptual contents to be incorporated in their alethic modal form. Even if it is properly objective_{sm} when a subject is committed to incompatible commitments in the sense that it is not up to anyone's attitudes immediately whether the commitments are incompatible or not, it does not immediately follow that the subject is in a state where she *ought to or may* do something, namely rectify the incompatibility.

Another basic reason the genuineness of commitments cannot fall out of an account that establishes their semantic grading as properly objective is that the latter possibility already presumes that some commitments are genuinely attributed. For only genuinely attributed commitments can be assessed according to their semantic correctness in the *primary* sense. The reason why the parrot cannot be said to *really* represent the property of redness by its squawking is that the parrot is not *really* undertaking any commitments, or undertakes commitments only in the *secondary*, derivative sense. But if that is true, the fact that some attributions of a human subject can be assessed as correct in the primary sense cannot be a *reason* to say that her attributions are genuine, for the (assumed) genuineness is a necessary precondition for the semantic assessment in the primary sense. I shall examine this line of argument more thoroughly below.

12.6 Motivating Pragmatic Objectivity on Independent Grounds

I have now argued why the reasons to support the pragmatic objectivity of normative force do not simply 'fall out' of the account of representational purport and success on the semantic side of objectivity, even though Brandom appears to claim the contrary at places. Next, I shall consider possible independent or complementary reasons in MIE for supporting the idea of pragmatic objectivity of normative force.

The problem can be usefully approached by comparing two sorts of descriptions in terms of their justifications and explanatory benefits, respectively. On the one hand, we have the non-discursive parrot, on the other, a discursive community of merely rational scorekeepers engaged in an ADP, who are capable of making assertions and implicit material inferences but not making any of these explicit in conditional claims. For the sake of argument, I shall assume that Brandom's original idea of layer-caked rationality works and that the level of implicit-only norms can be made coherent. Brandom claims that although an external interpreter may correctly employ normative vocabulary in the primary sense while describing the rational community exhibiting *original intentionality*, a zoologist describing the parrot's behaviour is entitled only to a derivative deployment of the same normative vocabulary. Two distinct problems ensue from this dichotomy.

First, what is it that *justifies* the discrimination of descriptions in terms of primary and secondary intentionality? Here, Brandom seems to rely on two main reasons. On the one hand, only the rational community can be interpreted as attributing normative statuses to each other independently of the external interpreter, which appears not to be the case with parrots. On the other hand, only the rational subjects are capable of giving and asking for reasons from each other, which arguably is not the case with the parrot. To 'give and ask' for reasons here means the ability to make assertions in specifically linguistic form, the paradigm of which are declarative utterances.

Let's admit the first point for the sake of argument. It is not obviously clear why the parrot's ability to draw implicit inferences in practice, in the primary sense, would be conditional on its ability to attribute this very ability to another organism. Brandom's answer here comes as an application of Dennett's (1971) proposal that the main function of intentional vocabulary is to predict and explain the behaviour of

organisms exhibiting a certain level of complexity. To say that the chess computer decided to sacrifice a Knight in order to protect the King, which implies attributing to the machine certain desires and beliefs, is a much more convenient way to explain its behaviour than resorting to describing its code would be. However, this precisely introduces the distinction between primary and secondary intentionality, for in order to understand the attribution of intentionality to the machine, we must take the intentionality of the attributer as a given and non-explicable by any further attribution. The reason why the ability to attribute the intentional stance to another organism is a necessary condition for exhibiting primary intentionality follows from the idea that intentional vocabulary has its home in interpretation of behaviour. If the parrot's ability to draw material inferences must be explained by recourse to an attribution of this ability to it by another organism capable of making material inferences, and if we aim to avoid lapsing into a regress of attributions, the primary sense of the ability to draw inferences and make assertions (the paradigm intentional moves) must be understood as conditional on the ability to attribute this ability to another organism, which is the mark of primary intentionality.⁶²

Brandom differs from Dennett's original proposal importantly in two respects. First, he does not understand 'interpretation' *primarily* in terms of prediction and explanation. Although it's true that adopting the intentional stance is useful (most of the time) in explaining and predicting the behaviour of certain complex organisms, the justification of adopting the intentional stance is not primarily instrumental for Brandom. Rather,

[w]hat follows immediately from the attribution of intentional states that amount to a reason for action is just that (ceteris paribus) the individual who has that reason *ought* to act in a certain way. This 'ought' is a *rational* ought—someone with those beliefs and those desires is rationally obliged or committed to act in a certain way. The significance of the states attributed is in the first instance a matter of the force of the better reason, rational force. (MIE, 56)

In other words, subjects adopt the intentional stance towards one another not primarily to explain and predict each other's behaviour, but to treat each other accountable to the 'rational force of better reason'. A consequence of this move is that Brandom does not need to make the 'substantive rationality assumption' that

⁶² It follows that only a community of organisms is able to display what Brandom calls 'original intentionality', i.e. primary intentionality, which only its members attribute to each other (MIE, 61).

Dennett does, which states that subjects usually are sensitive to the force of better reason, in order to make sense of adopting the intentional stance. As we already saw in Section 12.1, for Brandom, the primary sense of being 'bound' by reasons is to be correctly assessable by them, not of being sensitive to them.

A second important point on which Brandom differs from, or rather radicalises, Dennett is in how he understands the relation between the primary and secondary intentional systems. In order to be coherent, the proposal must understand primary intentionality, which implies the ability to adopt the intentional stance towards other subjects, as also intelligible only as an object of interpretation. At the same time, the distinction between primary and secondary intentionality must not collapse, for then the proposal would become incoherent. Now, John Searle (1983) has argued that there is no way to meet the two requirements at the same time because in order to attribute primary intentionality to an organism, we must, at some point, make sense of the intentional attitudes as they are in themselves, not as they are attributed from a further perspective. In counter to that, Brandom argues that if we allow some attributions of primary intentionality to be *implicit* as opposed to *explicit*, the 'stance stance' can be saved from incoherence (MIE, 60-61).

That sums up Brandom's argument for why the parrot cannot exhibit primary intentionality without the ability to attribute intentional states to others. He also has another, related reason to justify why the parrot can only be described by normative and intentional vocabulary in the secondary sense, which is that the parrot cannot be interpreted as *asserting* anything by its gawking even where its voicings share a shallow morphological or phonetic connection to typical human linguistic utterances. Because the parrot does not assert, any argument which tries to justify the description of its behaviour in primary intentional terms by appealing to the explanatory and predictive benefits, even where these benefits are substantial, is moot against Brandom, for as we saw, he privileges the assessability sense of bindingness over sensitivity. The instrumental justification of the primary intentional (or normative) idiom has no traction for the parrot because no matter how well its behaviour appears to conform to our belief-desire explanatory model, the parrot is never *correctly* held responsible for its behaviour.

To this answer, two critical questions must be raised. First, why is it that the ability to make *assertions* plays such a significant role in determining whether an organism can be held responsible in the primary discursive sense? Second, what is it really that the merely rational community does differently than the parrot, which justifies the external interpreter's primary deployment of the normative idiom in describing their sounds and gestures, i.e. why is the *linguistic* form to be privileged in making certain sounds and gestures count as assertions?

The first question is easier to answer. The reason assertion plays such a central role in Brandom's theory of intentionality is that it is the minimal *pragmatic* unit in which one can assume *responsibility for some propositionally contentful* commitment. Since responsibility is the key prescriptive term for Brandom, so must be the assertion on the pragmatic side. Indeed, I believe that Brandom's 'iron triangle of discursiveness' (assertion on the pragmatic, declarative sentence on the syntactic and proposition on the semantic side) has the form it does because of his reliance on responsibility as the dominant form, which normative force takes.

Onto the second question of why is it that assertions must take primarily linguistic form in declarative sentences. Nicholas Griffin (2018) has attempted to pry an answer from Brandom's works to this question. His critical argumentation starts from the judgement that as is, Brandom has no credible way to account for the intentionality of non-linguistic creatures, such as parrots, which leads to the incredible claim that they have no intentionality whatsoever. Further, all of Brandom's various attempts to mitigate the issue turn out to worsen the original problem. In the following, I agree with Griffin's assessment that Brandom indeed has a problem with animal intentionality that threatens to deadlock his account in undesirable dualism. On the other hand, I find Griffin's reading of Brandom's attempts at mitigation uncharitable and partially wrong.

To begin with, we can briefly see why Brandom's theoretical set-up leaves non-linguistic animals with no intentionality whatsoever. For one, in his later writings, Brandom distinguishes between 'practical' and 'discursive' intentionality, admitting that many non-linguistic animals displayed practical, feedback-governed 'directedness at objects' long before any human spoke a word (BSD, 178). However, after flagging it, the distinction is mostly left a placeholder in Brandom's works,

much like the earlier difference between merely 'sentient' and 'sapient' creatures is. Most importantly, since merely practical intentionality is not genuinely normative, it cannot have the primary status, which leads Griffin (I think correctly) to conjecture that even practical intentionality should be understood as explanatorily derivative in relation to original, discursive, linguistic intentionality (2018, 5530). Alternatively, practical intentionality might just collapse to the purely dispositional algorithmic model of explaining behaviour, which would mesh well with Brandom's perennial inclination to juxtapose parrots with thermostats. In either case, the results are clear: non-linguistic animals do not display primary intentionality in any sense. On the other hand, to say that they exhibit intentionality in the derivative sense is to say that they do not *really* display intentionality any more than our sounds and marks, which merely convey it.

Moving on, there are reasons to be unhappy with that sentiment. I agree with Griffin's exegetical claim, according to which Brandom's main argument for why parrots et al. only showcase derivative intentionality is that the conceptual contents of non-linguistic animal behaviour depend on our linguistic, normative practices for their specification. The idea appears explicitly albeit abbreviated in Brandom's response to Dennett (2010):

[N]othing non- or pre-linguistic creatures can do can underwrite attributing to them intentional states whose contents are specifiable by the declarative sentences of some language—say, English. Nothing the dog can do can warrant our characterizing what it believes is buried near the tree is a *bone*, or that it is a *tree* that it is buried near. Those concepts have their boundaries delimited by a network of inferences that relate them to other concepts. And what a merely practically intentional creature can do cannot be sufficiently articulated and normatively controlled in the right way as to warrant *literal* attribution of states whose contents are specifiable by the use of those of *our* concepts. (Brandom 2010b, 306)

As Griffin accurately observes, Brandom seems to think that attributing beliefs (i.e. propositional, conceptual contents) to a non-linguistic animal is necessarily normatively constrained *only* by *our* discursive practices and never by the animals' own behaviour. Insofar as the *source* of the constraints on interpretations in part determines what the contents are, the presumed *contents* of animal intentionality must be understood as derivative because the source of normative constraints is always on our side. The reason it is always on our side is that only we can *specify*, i.e. linguistically assert, the presumed contents. Because linguistic assertion is necessary

for specifying conceptual contents, and because animals do not know language, the implicit contents of their behaviour are wholly dependent on our interpretation, hence derivative.

The first problem Griffin sees with the answer is that generally, the inability to *specify* the content of a belief (or some other propositional, conceptually contentful state) does not entail that there is no belief or concepts to be specified *at all*.⁶³ Furthermore, Brandom's pragmatist inferentialism itself provides some arguments to the effect that specifying conceptual contents is less abstract and intellectualist and more a matter of practical engagement with worldly objects. The example of a litmus paper test extensively discussed in Section 11.5 is readily compared to a dog's practical engagement with a bone. In both cases, there is clearly a complex, feedback-governed algorithmic process at foot, and on the surface, it seems plausible to specify some, albeit vague and fallible, pattern of entirely implicit 'inferential' relations in the dog's behaviour. It is noteworthy that the dog's 'concept' of bone need not resemble in its inferential articulation that of any human's, for Brandom allows even massive differences in the mastery of the same concept between subjects (Griffin 2018, 5532).

The argument here is painted in broad strokes, but I think its pull is definitely against Brandom. For one, Brandom seems hard-pressed to deny that we can make *no* plausible conjectures about the content of the dog's concept of bone based on its behaviour with it. The counterargument must (and I think does) hit earlier, namely on the assumption that in order to posit a belief to an organism, it must be *in principle* possible to specify its content. Here, things become more complicated than I think Griffin implies, as his examples are somewhat shallow:

In the case of humans, we regularly attribute beliefs and other intentional states even when their content is unknown, as when we say, after some egregiously stupid performance, 'I have no idea what he was thinking'. The claim is not that he had no thoughts at all, but that we are at a loss to understand what they were. Or consider the participant in a magic show who is asked to think of a number. The audience attributes an intentional state to the participant, but is unable to specify its content. If we can do this with humans, it would need to be shown why we cannot do it with animals. (2018, 5531-5532)

⁶³ This intuitive idea is powerfully levied against Dennett's original, eliminativist stance stance by Alex Byrne (1998).

The analogy limps because in the case of humans with unknown beliefs, we can oftentimes always request them to relate their contents to us, which is not possible with non-linguistic animals. So, that sets a stark contrast to the senses in which a belief's content is 'unknown': the key point is in what sense the dog's allegedly unknown 'beliefs' are subject to the normative evaluation of correctness to begin with. Although our *method* of dissecting the practical inferential commitments of a dog engaged with a bone might, in relevant respects, greatly resemble those we use when observing an elaborate, tacit effort of a human to detach a car from a ditch (Griffin's example), the argumentative burden is primarily set on the *justification* of applying the method, which, according to Brandom, differs in principle in the two cases.

To examine the point, note first that Griffin, in my view, correctly takes Brandom's 'no-specifiable-content' claim as arguing for ontological results from epistemic premises. Roughly: 'Because we cannot know the contents of the dog's beliefs, we cannot say it has any beliefs at all' – a flagrant implausibility when applied to the human case. However, what Griffin seems to miss through his critical arguments is precisely that Brandom takes epistemic conditions as criteria for ontological postulation with (discursive) norms under the programmatic heading of phenomenalism about norms. An example:

At times [...] Brandom writes as if the presence of derivative intentionality in the brutes is a consequence or effect of the existence of linguistic practices among other animals which ascribe intentionality to them. But this can hardly be the case, and Brandom is surely not maintaining that it is. The mere making of the ascription is not sufficient to ensure that an object has the property ascribed; let alone that it have it as a result of the ascription. (2018, 5534)⁶⁴

For Brandom, intentionality is a normative status to be explained by normative attitudes, such as attributions of commitment. The source of normativity is in the attitudes, but only as they figure in linguistic practices or abilities. So, *contra* Griffin's complaint that this order of explanation either leaves animals with no real intentionality or then cooks up a story, according to which their derivative intentionality is awarded in the way of Stanley Kubrick's *2001: A Space Odyssey*'s monolith, I'd say that for Brandom, natural language just is the monolith, the proviso

⁶⁴ Dennett (2010) also relates his worries about this possible reading and quickly rejects it.

being that the apes not *only* found it, but *also* made it. So, perhaps they are also able to rent it to other animals from time to time.

What I mean by that is that if we are to take Brandom's institution of norms by attitudes claim seriously, and if we agree with him that genuine intentionality presumes linguistic, assertoric practices, it inevitably follows that the intentionality of non-linguistic animals somehow depends on our ability to ascribe that intentionality to them in the derivative sense. Of course, the mere ascription of a property cannot institute the properties themselves, Cambridge properties exempted. But Brandom's point precisely is that intentionality in any form is not a special sort of *property* as it was for Descartes, but a special sort of *propriety* as it was for Kant (MIE, 9). Oversight regarding this facet of Brandom's theory is why I am inclined to call Griffin's reading uncharitable and partially wrong.

Griffin does better, though he is not the first to do so, in pressing Brandom on the question of why it is *necessary* to describe the human or the proto-hominid behaviour in terms of scorekeeping, assertional proprieties, while unjustified to extend these descriptions in the primary sense to non-linguistic animals. The reason cannot simply be that animals do not assert because they do not know language, namely make declarative statements. For while it is largely uncontroversial that most animals that we intuitively count as exhibiting intentional behaviour do not have a language – whatever that means precisely – it is much less uncontroversial whether one can make assertions without being able to make declarative statements. (Note that it is more or less trivial that *humans* can assert without uttering a declarative sentence, say when I shake my head in response to a question.) What Brandom needs to explain is why the practical engagement of the dog with the world is nowhere *sufficient* to count as assertive, hence normative in the primary sense, while the proto-hominid community is *necessarily* describable in the scorekeeping, normative idiom (Griffin 2018, 5541).

Furthermore, resolving these problems still leaves Brandom with an entirely different hurdle to jump. For even if it is necessary to *describe* (or 'explain' as Griffin says) the proto-hominid community in the normative idiom, while it is not justified to do so in the primary sense with the dog and the parrot, nothing is yet said of what abilities are required to deploy propositional, conceptual contents implicitly in

practice. The claim that propositional contents have to be described in terms of linguistic, assertoric practices does not entail the claim that nothing can practically deploy propositional contents without being able to participate in linguistic assertoric practices – it just means that the dog cannot describe or explain what it is doing, not what it *is* what it is doing (Griffin 2018, 5542).

As far as I can see, Brandom's reasoning for why non-linguistic animal behaviour cannot be interpreted as involving implicit discursive norms is that we cannot correctly employ our scorekeeping on them, the reason being that:

1. The animal cannot be held responsible, because
2. we cannot say what it should be held responsible for, because
3. the animal does not give and ask for reasons for what it should be held responsible for.

The weakest link in the reasoning is arguably (3.). For, sure enough, being unable to *give* or *ask* for reasons does not entail that one *has* no reasons for one's behaviour. This is where Brandom takes back from epistemological externalism what he concedes to it elsewhere. On the one hand, Brandom admits that the archaeologist who can reliably perceptually discriminate between the Aztec and Toltec potsherds need not be able to articulate her reasons for being considered as entitled to her observational commitments, he thinks that *someone* needs to be able to articulate those reasons in order to attribute to her the normative status as a reliable observer (AR, 102).

As things stand, the score on pragmatic objectivity of normative force is this. The argument that genuine normative force 'falls out' from the argument for properly objective semantic contents does not work. The independent arguments that draw from Dennett's stance stance lead us to question premise (3) of the reasoning above, namely why is it that the ability to articulate reasons is necessary for the ability to have reasons.

Now, perhaps not much needs to be conceded to cash in the pragmatic objectivity of normative force. For it appears to suffice to drop the requirement that the scorekeeping model is *necessary* for instituting normative statuses, and that it is merely

sufficient to do so. This enables us to drop premise (3.) above and concede that non-linguistic animals too can display implicit propositional contents in their behaviour. These commitments are necessarily linked, for if it is necessary to employ the normative scorekeeping idiom in interpreting the merely rational community, it must also be unjustified to employ the same model for interpreting the behaviour of non-linguistic animals in the primary sense. What man must do, the brute cannot.

However, to downgrade the claim that it is merely *coherent* and not necessary to interpret the rational community as engaged in a norm-instituting practice turns out to have dire consequences for the genuine status of *our* discursive sovereignty, as I shall argue in the next section.

12.7 Appearance of Necessary Reasons

Above in Section 12.3, I presented the genuineness test, which Brandom's account of normative force must pass in order to be considered objective. The main requirement set by the test is that the community's attitudes must involve the possibility of global error regarding the force of the norms they bind themselves with, such that they could all be potentially in error whether a given (discursive) norm of theirs is really in force or not. No norm that fails this test can be *objectively* in force for the community members. My claim now is that the test as presented is too coarse-grained; considering the previous sections, we must introduce two further grades. The basic reason is analogous to the distinction, already encountered in the case of semantic objectivity in Sections 11.3 and 11.5, between the community members *taking* their own norms as objective_{sm} in the (PO) sense and the norms *really* being objective_{sm} in the (PO) sense. Likewise, in the case of force, we need to distinguish between

(AI-appearance)_{pr} A norm *n* *appears* attitude-immanent for community *C* iff it is not possible for everyone in *C* to *take* themselves as mistaken about the force of *n*.

(AI-reality)_{pr} A norm *n* *is* attitude-immanent for community *C* iff it is not possible for everyone in *C* to *be* mistaken about the force of *n*.

and, respectively

(AT-appearance_{pr}) A norm *n* appears attitude-transcendent for community *C* iff it is possible for everyone in *C* to *take* themselves as mistaken about the force of *n*.

(AT-reality_{pr}) A norm *n* is attitude-transcendent for community *C* iff it is possible for everyone in *C* to *be* mistaken about the force of *n*.

Here, it is good to remind ourselves of Brandom's methodological commitment to phenomenalism about norms, which states that the general course of explaining norms should begin with the phenomenal question of how norms *appear* to the community members in their practical interactions, and from there, aim to describe how they come to institute the *reality* of the norms themselves. Moreover, the more demanding conceptual (AT) norms are supposed to arise from the more primitive (AI) ones. So, just as *(AI-appearance_{pr})* is a necessary condition for *(AI-reality_{pr})*, *(AI-reality_{pr})* is a necessary condition for *(AT-appearance_{pr})*, which finally is a necessary condition for *(AT-reality_{pr})*, or the objectivity grading_{pr} which Brandom aims for under the Kantian thesis of the autonomy of reason.

I think it is sufficiently clear why neither *(AI-appearance_{pr})* nor *(AI-reality_{pr})* suffice for the Kantian autonomy of reason, for both fail to underscore the specifically global possibility of normative failure demanded by the genuineness test. However, a little further argumentation is needed to show why *(AT-appearance_{pr})* alone is not sufficient for the Kantian autonomy of reason, and why the stronger biconditional of *(AT-reality_{pr})* is required. Incidentally, the point is connected to the reason it must be *necessary* for the external interpreter to ascribe to the merely rational community genuine normative force in order for the external interpreter's own practice to count as exhibiting genuine normative force in the *(AT-reality_{pr})* sense, and why it must be necessarily unjustified to attribute intentional contents in the primary sense to non-linguistic animal behaviour.

To explain why this is, let us start by returning to the non-discursive proto-hominid community and combine it – this time, with the deeply religious community we met in Section 11.2. The external observer interprets the community as taking themselves

to be bound by the attitude-transcendent normative force in the form of their allegedly divine marriage code. At the same time, the external interpreter does not herself endorse the marriage code; thus, she does not take the natives to be really bound by the non-conceptual, (*AT-appearance_{pr}*) norms they all take themselves to be bound by. This is because he recognises nothing in the (practice-independent) world, which could determine the content of the marriage norms. But if the only *actual* source (according to the external interpreter) of the norms' determinacy is in the attitudes, then only the attitudes determine the content, hence force, if they do. In other words, the external observer differentiates between (*AT-appearance_{pr}*) and (*AT-reality_{pr}*) in case of the natives, thus taking it that although they all take themselves to institute genuine, non-conceptual marriage norms (both content and force), they do not *really* do so.

Now, the next step is to see that interpreting the community with non-conceptual marriage norms is not essentially different from interpreting them with their conceptual norms. In both cases, the external observer may differentiate between what the community members take themselves as being bound by and what, if anything, they are really bound by. Crucially, this interpretation is possible if it is not *necessary* for the external interpreter to describe the community's attitudes in genuinely normative terms, i.e. as what they *ought* or *may* do and not merely what they *take* each other as obliged and permitted to do.

Third, if the merely behaviouristic description (i.e. the description only of what the subjects take to be correct, not what is correct) of the community is possible for *all* norms the external interpreter ascribes to them, then it is impossible that the perspectives of the natives and the external interpreter would collapse in the sense required by the final move of MIE. For in the case where the external interpreter does not take the natives to be really bound by their own norms, he does not take them to really be engaged in the scorekeeping practice, which is what he presumably takes himself and his colleagues in the university to be doing.

The fourth, final move is to realise that *we ourselves* might be in the exact same relation to some alien interpreter as the external observer is to the proto-hominid or merely rational community. That is, it is possible that although we necessarily take ourselves as bound by genuine norms, we are, in fact, not really undertaking genuine

commitments as seen from some hypothetical alien interpreter's perspective, for whom our actions are sufficiently explainable in purely behaviouristic terms. But if that is the case, the Kantian autonomy of reason is not fulfilled, for we could not all *be* mistaken about the force of our norms since according to the alien interpreter, the norms are not genuinely binding for us to begin with: the possibility of an universal mistake is only an appearance, not reality, and the genuineness test has not really been passed.

It does not matter for the validity of this argument that there is no alien interpreter capable of either affirming or denying genuine normative practices from us. If it is possible for the hypothetical alien interpreter to deny genuine normativity from us, that is sufficient to show that we are, in fact, in the (*AT-appearance_{pr}*) category instead of (*AT-reality_{pr}*). But what argument would show that the hypothetical alien interpreter would *not* be forced by reasons to ascribe genuine norms to us? In response to McDowell (2005), who effectively raises the same concern from another angle,⁶⁵ Brandom complains that it would be implausible to the extreme not to interpret the practice as conferring propositional contents on expressions, and hence, as genuinely normative. He lists all the logical-expressivist structure of our language that he argues can be 'mapped on' the scorekeeping practice, such as conditionals, negation, language entry, exit moves, and so on, with the implication that if it walks like a duck, quacks like a duck, then surely enough:

If something is missing here, it is something *magical* — for this complaint is like that of the solipsist who insists that whatever *behavior* other humans might exhibit, it is nonetheless not accompanied by what *he* has: a *mind*. If showing the broadly inferential role of all of these locutions is not producing the pudding, what *could* count as doing so? (Brandom 2005, 240)

It is curious to me how Brandom's objection to McDowell's scepticism corresponds well with Griffin's objection to Brandom on the point of denying animals genuine intentionality. For what Griffin effectively argues is that it is implausible to the extreme (arcane, even) to say that a dog's behaviour with the bone is intentional only

⁶⁵ McDowell argues that the scorekeeping practice is not sufficient to confer propositional contents on expressions, which hence do not count as genuine assertions and inferences. But if the scorekeeping practice is not sufficient for this, it is not necessary for the alien (for McDowell, Martian) interpreter to ascribe genuine normative force to the practice; 'genuine' in the sense of (*AT-reality_{pr}*).

in the derivative as-if sense, just like McDowell argues that the scorekeeping practice could be interpreted by Martians in a similar as-if-genuinely-normative vein.⁶⁶

Indeed, I believe that the four cases of animals, us, the proto-hominids, and the Martians, are actually more than vaguely analogical; the two sets of relations are argumentatively linked, as I shall now argue.

The question that McDowell presents to Brandom is: why would it be necessary for the Martians to interpret the scorekeeping practice as genuinely normative? I have claimed that if it is possible for the Martians to coherently not to take scorekeeping as genuinely normative, it is also possible for them to deny *our* claim to genuine normativity under the assumption that we are, in fact, playing the scorekeeping practice. Observe now that the reasons the Martian has to deny our (assumedly scorekeepers) normativity are the same that the scorekeepers themselves have to deny animals primary intentionality. In both cases, what is ultimately denied is *discursive sovereignty*, the ability to make normative force really binding. It follows that the argument for why the Martians are *obliged by reason* to recognise our actual practices as genuinely normative must also serve as a reason for why it is *unjustified* to ascribe implicit norms and propositional contents to non-linguistic animals in the primary sense. And that requires showing why linguistic abilities, particularly the ability to give and ask for reasons, is necessary for having reasons to begin with.

12.7.1 Norms as Inner Reality: Response to Peregrin and Kiesselbach

Before continuing with my main line of argumentation, I think it is illuminating to refute a counterargument that might have arisen at this point, meant to mend the ails of genuine implicit normativity that certain authors sympathetic to Brandom have pursued on his behalf. I call the idea the 'inner reality' proposal, and its

⁶⁶ I agree with Wanderer (2008, 88-89) that Brandom's defence against McDowell is weak. The main reason is that in order to accept the implausibility of denying the scorekeeping practice genuine normativity, one also needs to accept the key aspects of Brandom's commitment to phenomenalism about norms, namely the Davidsonian idea that if a set of practices is interpretable by us as rational, the practice thereby just is rational, for rationality is a function of our interpretational practices.

representatives comprise Jaroslav Peregrin (2012; 2014) and Matthias Kiesselbach (2014; 2020).

To recap, the problem of implicit normativity that I have focused on in the preceding sections concerns the status of their force as *genuine*. What that means is that normative acts, such as attributions and undertakings of commitment must not be conceptually reducible to normative attitudes in the sense that whether a given commitment is attributed correctly or not does not collapse to whether *everyone takes* it to be correctly attributed or not. In short, genuinely normative acts must be capable of failure just like normal acts.

This criterion for the generically normativist order of explanation works only if the theorist also undertakes a commitment to the theoretical possibility of giving a broadly naturalistic socio-historical account of how the normative powers might have plausibly arisen in the course of history, or in other words, if the theorist denies that normative acts would be primitive in the socio-historical sense. In Section 6.1 I gave some independent reasons in the context of examining Ginsborg's proposal for why primitivism about discursive normativity is not an appealing option, although here, it suffices to note that Peregrin and Kiesselbach do not submit to primitivism in the socio-historical sense.

Note that the commitment to socio-historical non-primitivism does not entail that one must be able to *prove the truth* of any single such socio-historical account, which arguably is impossible due to the lack of sufficient empirical evidence. The threshold is lower in that one must be able to give a broadly plausible 'just-so story' of how the normative powers *could have* originated during the course of natural evolution and social history. A key part of such a story is explaining what would count as trying yet failing to institute genuine discursive norms on the global scale.⁶⁷

Peregrin's general strategy for making implicit norms intelligible in the broadly naturalistic socio-historical sense follows Sellar's idea of two distinct spaces: the

⁶⁷ Recall that although Brandomian normative acts must not be special in the sense of being immune to failure, they may still be special in other respects, e.g. by being capable of creating meaning retroactively in time; something Peregrin at least also seems to approve of (2014, 86-87).

ordinary external, natural, causal, and empirical space and the inner, normative, and conceptual space. Peregrin compares the latter to a kind of house:

Rules have an inner and an outer face. From the outside they, and the spaces they create, can be simply *described*: we can report on complicated linguistic practices that are going on within a community allowing members to use 'signals' to achieve complicated things. However, from the inside the spaces can be *inhabited*: we can *accept* the rules, making them into virtual 'walls' of a 'dwelling' we share with other people. Unlike a normal dwelling built from stone or wood, the walls of this one stand and fall with the attitudes of its dwellers. This creates the need for specific kinds of acts in order to support them. (2014, 89)

This is a homely enough metaphor since we understand perfectly well what it is to observe and describe a house from outside and what it is to live inside one. One can also perceive definite Heideggerian shades here, not only as concerns the metaphor of 'dwelling in a language', but also the idea that at bottom there is only one space, and language must be understood as integrated with the ordinary one in some fashion. The general fundamental pragmatist idea, to which Peregrin subscribes, is that the integration must be understood in terms of certain special actions and doings – the normative actions, to be specific.

The metaphor is telling in the additional sense that animals do not live in houses (only in dens, nests, or burrows) and that we put up walls partly in order to keep them out. This is not merely an unwarranted snarky remark, but gestures to how well the idea that natural language has a normative deep-structure in the form of norms implicit in practice can explain the leverage problem mentioned in Section 9.5. Peregrin's discussion (in 2014, Ch.6) of the evolution of language, together with Sellar's ideas of pattern-governed behaviour, offers a promising shape of the solution in terms of implicit rule-following that I think works quite well with Brandom's program.

However, the problem is that the reason the 'inner space' proposal works well in answering the leverage problem is that it is ultimately closer to the Platonist-Cartesian position regarding the distribution of true intentionality in nature, and thus faces its real difficulties in answering the emergence problem. We can see this by examining Peregrin's characterisation of primary normative actions, which he calls 'normatives':

[Normatives] do not report that something is the case, they point out that something *ought to be the case*, and hence they always involve the utterer's taking a rule for being in force, her endorsing it. In this respect, they are similar to oaths of loyalty: they always involve one's decision to assume a certain status, namely to bind oneself by a rule, and in this sense they *institute* something (a certain social link) rather than *report* it. However, the case when the institution happens in a single instant (as in the case of signing an oath) is only a very special case; more generally, binding oneself with a rule is more like the case of loyalty that is not formally established with an instant oath, but is continuously testified by one's performances and declarations. Normatives of this kind involve instituting and upholding – or, as the case may be, amending or contravening – a rule. (Peregrin 2014, 77; cf. Peregrin 2012)

This is very close to Brandom's official stance in MIE, down to the vocabulary of 'institution'. Like Brandom, Peregrin also thinks that normative attitudes, which normatives make explicit, can be adequately described in their external reality by the means of, e.g. theories of cognitive science. He mentions Gendler's (2008) 'aliefs' and Dreyfuss's (1999) 'absorbed copings' as speculatively fruitful accounts. Yet as to describing the attitudes in the inner, normative sense of 'holding-correct' (or 'holding-true') Peregrin appeals to primitivism, saying that they are not explainable in any simpler terms.

How is the appeal to primitivism to be negotiated with the task, which Peregrin clearly sees this to be important, of giving a broadly naturalistic socio-historical answer to the emergence problem of intentionality? This is where the appeal to internal reality becomes important. According to Peregrin, the normative force is primitive only in the internal sense, so that it does not clash with the external explanation, as well as the description of our linguistic practices and normative attitudes. We cannot conceptually analyse the concept of 'holding-correct' (or 'holding-true') by some more primitive kind of vocabulary. The function of normative vocabulary is primarily prescriptive (which, in my terminology, covers proscriptive and permissive normativity) in the internal sense, which means that community members perceiving the violations of norms are disposed to sanction the offender – in an extreme case, by rejecting him from the community completely, just like the shepherd boy in Aesop's fable. Although there is no obligation to belong to any community, if one belongs in one, then one is legitimately (correctly) taken as subject to the sanctions of other practitioners.

Peregrin's thinking seems to dovetail closely with Brandom's, but on closer inspection, it becomes clear that the 'inner reality' option cannot be Brandom's where he is committed to the synthesising solution between phenomenalism about norms and normative phenomenalism. For if it is possible to describe and explain linguistic practice and normative attitudes coherently *only* externally, then some alien interpreters would not be forced by reasons to recognise the scorekeeping practice as genuinely normative. It then follows that either our linguistic practice is not *explainable* in terms of scorekeeping, or that scorekeeping is not genuinely normative, both of which mean the failure of the synthesis in different senses, as was explained in Sections 10.1 and 10.2.

The key point where Peregrin's inner reality proposal breaks with Brandom's synthesis can be articulated in the vocabulary of (\mathcal{AT} -*appearance_{pr}*) and (\mathcal{AT} -*reality_{pr}*) presented in Section 12.7. What Peregrin effectively claims is that normative actions and force can be genuine under (\mathcal{AT} -*appearance_{pr}*), i.e. in the scenario where the participants within the practice (necessarily) take the norms to be in force for each other, but where the norms are not in force for anyone, as seen from the external, broadly behaviouristic perspective. But I already argued that (\mathcal{AT} -*appearance_{pr}*) is not sufficient for genuine normative force. Briefly, it is not sufficient for the same basic reason (\mathcal{AT}_{sm}) is not sufficient for (\mathcal{PO}_{sm}): because the assertion that our claims appear to purport to represent the world does not entail that they actually do. Similarly, from the fact that genuine norms (necessarily) appear to be in force for us, it does not follow that they actually are. It is because of this crucial difference that Peregrin cannot explain what it would mean for normative actions to fail in the global sense, or what more is needed to step away from talk of force's (in)corrigible appearance to its (corrigible) existence.⁶⁸

68 It should be noted that in his later writings, Peregrin (2021) has, in fact, been explicit about the important differences of his views in relation to Brandom. For one, Peregrin gives up Brandom's idea that the normative, conceptual space would be closed in on itself, i.e. there would be something real there which no non-normative external description could capture. The distinction between descriptive and normative speech acts must not be interpreted in ontological or epistemic, but rather in phenomenological terms: 'It is one thing to record that something is (treated as) a rule in a community and that some items are (treated as) correct/incorrect according to it; and it is another thing to endorse the rule and treat the items as correct' (2021, 125). Second, Peregrin later holds that some normative attitudes, e.g. accepting the modus ponens rule, are basic in the sense that they cannot be themselves evaluated by further normative attitudes (at least by humans). It is not perfectly clear (to me) to what extent these positions are present already in the work I cited above; here, my focus is not in finding

In fact, I think that, already in MIE, Brandom denies from himself the possibility of endorsing the inner reality solution advanced by Peregrin. For in that work, Brandom commits himself to the possibility of formulating a theory of *canonical designators* for discursive norms (MIE, 626, fn.). Briefly, canonical designators form a kind of privileged set of descriptive terms, the substitutability for any other descriptive term of some ontological category in the context of a claim determines whether the entity denoted by the term exists in the ontological category or not. For example, according to Brandom, the class of canonical designators for macro-physically descriptive terms, such as 'horse', are spatiotemporal coordinates such that a continuous spatiotemporal trajectory can be traced from the speaker's egocentric space to the region occupied by the entity denoted by the term used. Put less clumsily, to claim that an entity exists in the macro-physical sense is to claim that it exists or has existed in some region of space-time accessible to the speaker in principle, where the spatiotemporal region coordinates serve as a canonical designation of the macro-physical ontic category (MIE, 444).

Brandom suggests that different ontic categories, such as numbers, fictional entities, and so on, come with different senses of existence respectively; thus, different sets of canonical designation are available to them. He also mentions that it is possible to have many contenders for one's choice of a canonical designation set in a given ontic category, although I must assume it to be implicit that every proper ontic category has only one *correct* such set. That is to say, unless two canonical designation sets produce exactly the same ontic values (exists or not) for terms of a certain category, the two must be incompatible with each other – and at most, one may be correct. The reason why both cannot be correct is that it would mean one and the same ontic category could have distinct yet compatible senses of existence for its entities, which is hard to make sense of considering how the term 'existence' is usually used, namely that two entities of the same ontic category must exist (if they do) in the same sense.⁶⁹

out the truth of the matter, nor in discussing Peregrin's otherwise interesting proposals. The important point is that these answers cannot be part of Brandom's official project.

⁶⁹ This is compatible with saying that for every ontic category, there is only one correct set of canonical designators, yet for some, we will never reach an epistemic point where the issue is settled beyond reasonable doubt. For example, the fictional category could well be an example. Here, I am not claiming to have set an argument in stone regarding the interpretation of the predicate 'exists', but

Although Brandom has nowhere cashed out his commitment to proposing a set of canonical designators for discursive norms, if my interpretation above holds, then it follows that Peregrin's internal reality proposal cannot be Brandom's. If we can claim that discursive norms exist in the scorekeeping vocabulary, they do not merely appear to exist. And if they exist for us, they must exist also for the alien interpreters in the sense that they would thereby have a necessary reason to recognise the scorekeeping practice as genuinely normative. That is precisely where Peregrin's proposal breaks off with Brandom's.

Next, I shall examine another closely related proposal by Kiesselbach (2014). To start off, while his theory of 'linguistic calibration' is advanced as an independent theory from Brandom's, it is clear that the two approaches share many basic presumptions and methodological commitments as Kiesselbach himself acknowledges (2014, 435). Especially clear is his endorsement of the Sellarsian division between two spaces:

My claim, now, is that the central locus of normative vocabulary, crucially all talk of 'constitutive' or 'normative standards,' is within just such a calibration game, and that it is also within such a calibration game—i.e., from the perspective of a participant in it—that we must understand what it means to say that a norm is 'in force.' Our saying that a norm is in force or that some player is appropriately assessed on the basis of some standard ('has some reason,' in the vernacular) reflects our current stance within a calibration game—whether we are theorists, or other (more ordinary) participants in the calibration practice. And this, in principle, is all that needs to be said in an account of how constitutive norms come into force. (Kiesselbach 2014, 437)⁷⁰

The idea is close to Peregrin's in that normative acts and force appear only in the internal space or within the calibration game, not externally. While this means that norms cannot be strictly speaking described, only avowed, their force is still genuine in the sense that subjects within the practices can be correctly criticised on their basis even when they are not particularly sensitive to them. As for Brandom, as well as Kiesselbach, the primary sense in which one is subject to normative force is

rather only pointing out the reasonable consequences of what seems like the most natural reading of Brandom's thought to me.

70 By 'calibration game', Kiesselbach means a linguistic practice, analogical to Brandom's scorekeeping, where the participants evaluate each other's moves and seek to align each other's standards of evaluation in order to achieve communication. The details of linguistic calibration are not relevant in this context.

(legitimate) assessability by some standard, not sensitivity to being guided or motivated by it.

In my work (2020), I criticised both Kiesselbach and Peregrin (and hence, Brandom indirectly) for not explaining what the 'internal' space amounts to as a theory of how discursive sovereignty can be genuine. In particular, how is it that subjects can be *correctly* held responsible in view of discursive norms under phenomenalism about norms, which states that only attitudes are sufficient and necessary to make norms binding? For as I have explained, it follows that to be correctly held as subject to a norm means to be taken as something properly subject to a norm. Either some attitudes (or their totality) must be considered incorrigible in some takings, or then the account lapses into an infinite regress where no norms can be said to be correctly enforced, and no genuine discursive sovereignty is instituted.⁷¹

In his response, Kiesselbach (2020) appears to think that a correct assessment by discursive norms can be understood as successful communication. The internal space formed by intersubjective efforts of the participants to interpret one another very much resembles Davidson's idea of interpretive charity. One of the key statements that Davidson pursued under the idea of charitable interpretation was that successful communication need not presuppose 'shared language', by which he meant that the meanings of words need not be the same for the speaker and the hearer *prior* to their understanding each other. It suffices that the hearer can make a passing interpretation on the go of what the speaker says in the hearer's own words. Even an ordinary conversation between two speakers of, say, 'English' is indeed a case of translating subject-relative idioms, according to Davidson (Davidson, 1986b).

⁷¹Observe here that it does not help to say that there is no 'totality of attitudes', following Brandom's historical, three-judges' model of conceptual determination, which shows that the correctness of applications is temporally open both to the past and future. So, strictly speaking, the community is never 'finished' but may always in principle allow for new members to join in at some later point, or perhaps even include some members retroactively. This is, I think, closely analogical to Fossen's defence of Brandom against Pippin, discussed in Section 11.2, where the idea was that even non-conceptual norms can be objective in the sense of being contested in the community. But contestability does not entail corrigibility of the totality of attitudes in the counterfactual sense that if everyone agreed which marriage norms are in force, then the community could not be wrong about the matter. Similarly, *if* every (potential) judge in the historical succession *were* to agree in perpetuity that a given norm was correctly taken to be in force, no genuineness test for the reality of norms has been passed unless they could also *be* wrong, and not simply possibly *taken* to be wrong, by some further judge's judgement. In short, contestability (lack of the totality of agreeing attitudes) does not entail corrigibility in the counterfactual sense, which is the relevant sense for passing the genuineness test.

The way how Kiesselbach thinks charity can solve the threatening regress discussed above is that to be correctly taken as something subject to discursive norms is always a *local*, *temporary*, and *fallible* (my words, not his) act by the interpreter. It is local in that while the interpreter may successfully interpret some assertions by the speaker in her own words, it is rarely if ever possible to give an interpretation of all their assertions. It is temporary in that while two subjects may not be able to understand each other at t_1 , they may reach an understanding at t_2 , and they may also later lose a previously gained ability to understand each other. And communication is fallible in that although two subjects may think they are successfully interpreting each other's assertions, they do not really do so. Because linguistic interaction is always in such flux and flow, uncertainty about the interpretation of the speaker as something subject to some discursive norms does not matter. Hence, to point to a regress of interpretations as to what really counts as correctly taking the speaker as something subject to norms simply reflects the messiness and feedback-governance of actual linguistic interaction rather than a foundational flaw in the model. So long as the participants can understand each other some of the time under some conditions, it makes sense to say that they bind each other with genuine discursive norms in the internal space of reasons (Kiesselbach 2020).

The core problem with this picture is that it conflates success in linguistic communication with the institution of genuine discursive normative force. Again, the basic point can be made by using the terminology introduced in Section 12.7. It might well be true that for me to successfully interpret your speech, I must necessarily treat both of us as bound by certain discursive norms, which is to affirm ($\langle AT\text{-}appearance_{pr} \rangle$). But it does not follow that by understanding each other, we have thereby 'instituted' genuine discursive normative force that exists in the sense of ($\langle AT\text{-}reality_{pr} \rangle$). In particular, even if it is true that *we* necessarily must treat each other as responsible for discursive norms in order to understand each other, it does not follow that an alien interpreter would likewise have to take our practice as genuinely normative, which, as already shown, entails the impossibility of global error and failure to pass the genuineness test. The internal space proposal cannot save genuine normativity precisely because genuine normative force requires assertability in the external (or the 'sideways-on', as Kiesselbach puts it) perspective.

Furthermore, it is questionable to say the least whether *we* necessarily must take ourselves as *responsible* to discursive norms in order to understand each other, i.e. whether even (*AT-appearance_{pr}*) is true of us. Davidson himself, for one, strongly doubted this:

Suppose someone learns to talk as others do, but feels no obligation whatever to do so. For this speaker obligation doesn't enter into it. We ask why she talks as others do. 'I don't do it because I think I should', she replies, 'I just do talk that way. I don't think I have an obligation to walk upright, it just comes naturally.' If what she says is true, would she not be speaking a language, or would she cease to be intelligible? In other words, what magic ingredient does holding oneself responsible to the usual way of speaking add to the usual way of speaking? (Davidson 2005, 117)

It is absurd to be obligated to a language; so far as the point of language is concerned, our only *obligation*, if that is the right word, is to speak in such a way as to accomplish our purpose by being understood as we expect and intend. It is an accident, though a likely one, if this requires that we speak as others in community do. (Davidson 2005, 118)

What Davidson argues is that *feeling responsible* for (or being sensitive to) discursive norms is not necessary for successful linguistic communication. While it is true that Kiesselbach, Brandom, and Peregrin privilege assessability over sensitivity, it is easy enough to extend Davidson's line of thinking by asking whether it is really *responsibility* that we attribute to the speaker while interpreting her. Davidson could have equally well formulated his point in this way too: 'It just comes to me naturally to interpret other people's speech in the way I'm inclined to understand the meaning of words; but they're not responsible to my (understanding of the) standards.' Would a hearer who did not treat her peers as *responsible* to meaning itself be unintelligible to them?

Finally, what comes to Kiesselbach's analysis of the background philosophical motivations implicit in my work (2020), I think he is well on the right track. An important joint in the controversy between normativists and anti-normativists reflecting a broader chasm in current analytic philosophy concerns the nature of reasons for action, of which linguistic action is one large subregion. Must we understand *meanings/norms themselves* as somehow contributing to the reasons for linguistic action, e.g. to the business of applying a word in one way rather than another? Or can we rather follow the 'Humean' route and understand (linguistic) action purely in causal, non-normative terms, e.g. by the belief-desire model?

To offer a kind of a retrospective specification of my views, while I think that linguistic action is usefully viewed as a causal, feedback-governed process, I am sceptical about the usefulness of the belief-desire model for describing action even in abstract principle. One reason I have discussed in Chapter 1 is that the whole notion of propositional content that is essentially implied by the belief-desire model becomes suspicious once we draw certain fairly modest conclusions from the Kripkenstein sceptical challenge. In my (2020) paper, I was happy enough to accept the Humean framework as a contrast for the normative approach. At the time, I had nothing better to offer and nevertheless, Glüer and Pagin, whose side I was holding, relied heavily on the belief-desire model. Their strongest point was, I felt, that they could do without the susceptible notion of 'genuine normativity', even if they, at the same time, leaned too hard on 'propositional contents' as a given.

12.8 Turner and the Disenchantment of Reason

At this point, an interesting puzzle ensues from the insistence that I have been tracing, namely that taking norms to be in force does not entail their being in force. However, *if it is necessary* for us (assumedly scorekeepers) to take some norms to be in force in order to make assertions and inferences, then it follows that we cannot assert that there *are no norms*. This is because to assert that one would have to *ex hypothesi* presuppose taking it that some discursive norms are in force, which would entail a pragmatic self-contradiction. So, we effectively end up in another one of Wittgenstein's paradoxes, namely whereof one cannot speak, one thereof must be silent.⁷²

In this section, I aim to critically examine the idea that we must necessarily take discursive norms to be in force if assertion, inference, and hence propositional contents are possible for us, or in other words, that (*AT-appearance_{pr}*) is true of us. I shall show that although there is an interesting kind of *perspectival effect* involved in a range of somewhat parallel cases, there are some good reasons to believe that the effect is no evidence of genuine normative force. Furthermore, I shall argue that even if Wittgensteinian silence holds for discursive normativity (i.e. it is impossible

⁷² The idea is made perhaps more explicitly already in (*Tractatus*, 5.61).

to assert either that norms exist or that they do not exist), the consequence is not necessarily a compromise quietism or renegeing back into normative primitivism, for the 'perspectival effect' can have a purely natural explanation in case of implicit norms, namely that they do not exist.

To begin with, it is necessary to further characterise the perspectival effect in more detail. This can be done by examining certain analogical cases, namely Edmund Husserl's criticism of logical naturalism and Hilary Putnam's criticism of cultural relativism. My brief elaboration of these cases draws heavily from Leila Haaparanta's (2019) discussion.

Husserl, along with Frege, sought to defend the normative autonomy of logic in relation to contingent human psychology. In particular, for Husserl, the validity and hence justification of logical laws and inferences were not due to the natural psychology of human subjects. Husserl's perhaps strongest argument, according to Haaparanta, is the following:

Logical naturalism claims that it is logically possible for there to be alternative logical systems or inferences in the sense that their validity is not due to contingent human psychology, but rather to some other form of psychology.

In claiming this, logical naturalism must apply the concept of logical possibility.

So, logical naturalism must *take* logical possibility to be valid in order to claim that it is *not* valid. (Haaparanta 2019, 188)

The contradiction here is *pragmatic* and not semantic in nature because what logical naturalism effectively tries to *say* contradicts the *saying* of it. Somewhat similarly as saying 'I am a humble person' is not something a truly humble person *would* say, to claim that logical possibility is not universally valid is not something that *can* be said while holding onto the very notion of logical possibility as it *appears* to contingent human psychology, namely as universally valid.

Another example of the perspectival effect comes from Hilary Putnam's criticism of cultural relativism. According to Haaparanta, Putnam argues in the following way:

Cultural relativist asserts that 'When A says that 'Snow is white', what she means is (whether she knows it or not) that snow is white in the sense determined by the norms of his cultural community'.

In asserting this, the cultural relativist, in order to be coherent, must understand his sentence 'Snow is white' as meaning 'Snow is white in the sense determined by the norms of my [the cultural relativist's] cultural community'.

When we embed the paraphrased claim in (2) to the original claim in (1), we get: 'When A says that 'snow is white', what she means is (whether she knows it or not) that according to the norms determined by A's cultural community as determined by the norms of my cultural community, snow is white'. (Haaparanta 2019, 189)

The important point is that (3) is not something that the cultural relativist can assert to hold *from the other culture's perspective* in the way of (1) because the other's culture's perspective is something she must construe for *every* sentence based on her own cultural norms, which are always *taken as given*. Again, this is not a semantic but a pragmatic self-contradiction: the naturalist and the relativist try to make assertions outside their own language, which fails not because there are no sentences outside their language, but because they themselves cannot step outside 'their own language', i.e. the language they happen to be *using*. For the same reason, notes Haaparanta, Husserl cannot assert, e.g. that 'There is only one possible logic', for that would mean 'stepping outside' the one he is stuck with (2019, 191).

It is well beyond my possibilities to examine what exactly is going on in this family of cases, or what the 'pragmatic perspectival effect' is really about. For my purposes, it suffices to recognise that it is a *real* effect; it is real because we can *see* that it is. The argument that I wish to advance in this section is that implicit normativity is *not* among the family of examples I have described as falling under the 'pragmatic perspectival effect'. To show this, it suffices to show that we do not necessarily need to assume norms to be in force in order to interpret linguistic behaviour as meaningful.

Turner (2010) has two broad strategies to argue against the necessity of assuming normativity virtually in any form in explaining human behaviour, including linguistic behaviour. The first strategy focuses on the internal problems of the normativist account, such as circularity and question-beggingness, while the second offers a positive, alternative account of how to explain some of the key phenomena that normativism takes as its strongest ground, namely the interpretation of linguistic behaviour.

The key, faulty move of the normativist, according to the first strategy, is to assert the primacy of a particular description. The basic idea of privileging the normative description of some key term is that it is the logically only way to talk about the phenomenon captured by the term, and that any alternative non-normative (e.g. causal) interpretation amounts to changing the topic. The problem here is to justify why certain vocabulary is alone logically necessary to describe what assumedly is a real phenomenon, such as reason, the ability to make material inferences and so on. Turner does excellent work in unearthing the inevitable circular reasoning, which the privileged description account is driven, e.g. in the case of Hans Kelsen's efforts to found the normative autonomy of legal vocabulary in relation to sociological one (2010, Ch.3).

While normativists of the Brandomian expressivist persuasion acknowledge the possibility of an external, non-normative description of the key phenomena, here too there persists an insistence that the expressive process itself cannot be sufficiently described in non-normative vocabulary without changing the topic. This is what Brandom does when he claims, without further argument, that in attributing intentional states, one necessarily attributes the '*rational ought*' (MIE, 56, *et passim*), or when Peregrin says that '*some kind of ought*' to is constitutive of the very meaning of *correct* (2014, 78-80).⁷³ Here, Turner is unable to catch Brandom with a blatant circularity: Brandom is right that it is coherent to describe the scorekeeping community's behaviour only in terms of internal sanctions and other normative terms. Turner's crucial objection is that it is not *necessary* to describe linguistic behaviour ('the expressive process') in normative terms. This suffices not only to show that the scorekeeping model does not pass the strong genuineness test, but also that we are not, in fact, stuck with the aftermath of the perspectival effect. In other words, Turner can show that neither (*AT-appearance_{pr}*) nor (*AT-reality_{pr}*) is true of us.

Turner's positive critical argument that hits Brandom's brand of normativism in particular draws from Davidson's philosophy of charity and communication. The pre-theoretical phenomenon, which both charity and the scorekeeping model are tasked to explain, is the hearer's ability to distinguish in her interpretation of the

⁷³ The exegetical attribution of this commitment to Peregrin is more complicated, as I mentioned in the previous section.

speaker between word-meaning and speaker-meaning. The ability to distinguish these two 'kinds' of meanings is central to understanding, for without it, the speaker cannot distinguish between what the speaker *actually* says (i.e. what is the literal meaning of his sentence) and what he *intends* to say while making some kind of error on the way.

Why is the ability to detect *errors* important for communication? The reasoning here starts with the premise that there is *de facto* usually some, often considerable variation in different interlocutors' dispositions to understand the meaning of words. If the only tool the interlocutors have for understanding each other is to attribute *different* meanings to the other's use in detecting a mismatch, the *convergence* of the interlocutor's usage-dispositions is impossible. Every time the hearer detects the speaker using a word in a way which she herself would consider being erroneous in her own case, if she thereby never seeks to correct the speaker but rather changes her interpretation of him, the usage-dispositions will keep diverging from each other until understanding becomes impossible.

Likewise, if understanding is to happen, then it is impossible for the hearer to *only* attribute erroneous usages to the speaker. The reason here is that attributing an error is only possible against the background of assuming an agreement in a broad swath of beliefs between the speaker and the hearer. Consider an example where we have made some progress in a radical translation of a speaker. Fortunately, the subject's idiolect mostly resembles English, i.e. the language we are disposed to interpret by our prior theory, yet many of her words continue to elude us regarding their meaning. One such case comes up as the subject utters the words 'That's a nice gavagai' while pointing to a running hare. This puzzles our prior theory interpretation, which had come to expect 'gavagi' to mean *rabbit*. We thus face a choice: assuming that the speaker knows the meaning of 'gavagi', either we attribute to him a false belief (as to the identity of the properties of hares and rabbits, the species of the particular animal pointed to, or both) or else we change our prior theory and hence interpret 'gavagai' to mean, most plausibly, *rabbit or hare*.

The point is that attributing an erroneous application to someone presumes attributing an indefinite but significantly vaster set of beliefs held true by both the interpreter and the interpretee. For example, suppose the interpreter chooses to

attribute to the speaker an erroneous use of 'gavagai', which (only) means *rabbit*. But that alone is not yet sufficient for the assertion 'That's a nice gavagai' to be incorrect in the context, for there are indefinite background beliefs, the presence of which makes A's assertion correct, e.g. that an animal running changes its species, or that the date is relevant for determining its species, etc. The interpreter can only attribute an erroneous use, and hence a false belief, against a vast background of beliefs held true by them both. Nothing, of course, stops attributing a massive falsehood in belief to the subject in the abstract, but in that case, it becomes impossible for the interpreter to understand the subject's speech and language. The vaster the discrepancy in the background beliefs that the interpreter attributes between himself and the subject, the smaller the set of instances of applying words where he can make the distinction between differing and erring usage. But if he cannot make this distinction at all, he cannot interpret the subject, because then he cannot make sense of what the subject's words mean and what the subject intends to mean by them, i.e. the distinction between speaker-meaning and word-meaning.

The lesson here is that understanding the speaker requires the ability to attribute both *different* and *erroneous* usages. The question then becomes: how is the speaker to choose between these in context? The principle of charity states that, assuming her goal is to understand the speaker, the hearer should always opt for the interpretation, which maximises the coherence of the combined collection of meanings, use, and beliefs that she attributes to him.

In contrast, according to scorekeeping, the hearer does not and should not aim for maximal coherence of interpretation, but rather simply to what is the *correct* interpretation of the expression used by the speaker. In other words, the scorekeeper holds the speaker responsible for the word-meanings (as she understands them) regardless of how irrational this renders the speaker. The minimal threshold of coherence that the scorekeeper allows to ground his interpretation is what is required to understand the speaker as someone rational or capable of speech at all.

Now, both models of communication as interpretation agree that the hearer must be able to take *some* utterances of the speaker as *erroneous* in order to understand him. The crucial difference is whether the 'error' must be given a normative interpretation or not – normative in the sense that the hearer treats the speaker as *responsible* (and/or

authoritative) to correct the error. In the previous section, I already offered Davidson's own, paraphrased answer, which doubted the necessity to give errors a normative interpretation in the sense of responsibility. Moreover, Turner's arguments put serious stress on the idea that the normative description is logically privileged. Yet, Brandom is committed to the idea that in order to understand the speaker's speech at all, the hearer must treat him responsible in the specifically rational, non-instrumental sense.⁷⁴

The ultimate critical claim that Turner advances against Brandom is that regarding explaining successful communication and how subjects can in practice distinguish between word-meaning and speaker-meaning, it is *metaphysically unnecessary* to presume genuine normative force as implicitly instituted in practices. The explanation can manage with far less, e.g. by empathic reasoning and the ability to simulate others' thinking. What the alternative, causal-only story about successful linguistic communication consists of precisely as such is irrelevant to the argument that the genuinely normative option is not forced upon us by any metaphysical concern.⁷⁵ This is the sense in which Turner, following Weber, can disenchant Kantian autonomous reason: by showing that it is not metaphysically posed upon the task of explaining human behaviour, including linguistic behaviour. The only way in which genuine normativity can squeeze a foot between the theoretical frame and the door is by assuming the need for a transcendental argument. The case here is analogical to what Shapiro argued on the semantic side back in Section 10.2: with the *need* to give a non-deflationary account of representation and propositional contents goes the *possibility* of giving one.

If Turner's arguments have the last word, then the quest for implicitly instituted normativity is done for. But Brandom yet has one big ace in his sleeve: the titanic *A Spirit of Trust*, to which we must turn our attention next.

⁷⁴ Turner (2010, 161) observes that error can indeed be understood as 'normative' for Davidson as well – but importantly, in a different sense than for Brandom. The issues here are intricate, but the main point is that normativity in the sense of Davidsonian charity is independent from external bindingness in Brandom's sense because the former notion does not presume that norms or rules are in force.

⁷⁵ It is true that there are good methodological reasons for why we have to presume the validity of some normative theory. For example, without presuming an ideal theory of reasoning, we cannot make sense of actual reasoning errors and bias. But the methodological necessity of the normative theory is not an argument for the metaphysical institution claim (Turner 2010, 166).

CHAPTER III: THE GEIST IN THE MACHINE

13 INTRODUCTION

To summarise the findings of the previous chapter, the main problem for Brandom's work downstream of MIE is to explain how discursive normative force can become genuinely binding on subjects, or how a community capable of original intentionality can sovereignly 'institute' discursive norms. While many of Brandom's critics take the explanation of *semantic* objectivity to be the crucial, ultimately unresolved problematic part of MIE, I argued that Brandom is, in fact, able to account for proper objectivity in principle already in MIE, supposing that problems relating to conceptual realism can be resolved. Instead, I followed the line emphasised by authors, such as Pippin (2005), Dennett (2010), and Turner (2010), in arguing that the crucial issues with Brandom's overall project concern the genuineness of normative force rather than the objectivity of semantic content. Furthermore, I specified and motivated the exact problems in Sections 12.3, 12.4, and 12.7.

My purpose in this chapter is to, first, bring the criticism of Chapter II to a conclusion by extending it to Brandom's most recent grand attempt at synthesising his founding commitments to normative phenomenalism and phenomenalism about norms under the theoretical gravity of ST. Second, I shall draw the results of the extended criticism into dialogue with the findings of Chapter I and the lessons of the sceptical challenge, the most important of which are pertinent to the temporal nature of meaning and the nature of metasemantic theorising.

This brief introduction serves two purposes. The first is to roughly summarise the main research problem of ST, as well as Brandom's proposed strategy for resolving it. The second is to summarise my argument as to what is wrong with, not just the solution, but with the problem as well.

As I have repeated several times, I think that a major (indeed, perhaps the most important) theoretical motive for Brandom is to solve the problem of reason's sovereignty by synthesising his founding commitments of phenomenalism about

norms and normative phenomenalism. I think that ST's key aim, that of solving Kant's Third Antinomy – 'a challenge to integrate reasons and causes' (ST, 558) – bears a close resemblance to this purpose, although it is true that ST's total aims transcend even that massive task. Described in these terms, the challenge is to explain how our meanings (conceptual, propositional contents) can be *constrained* in their use by more than merely causal factors, i.e. the attitudes considered as dispositional states, assuming that merely causal factors are insufficient to account for determinate contents. How is it that an external standard of semantic correctness can arise from the internal workings of the scorekeeping practice?

In the previous chapter, I already foreshadowed two of Brandom's main pieces for filling the puzzle. The first is to synthesise the Queen's Shilling and Kantian Autonomy models into a Hegelian mutual recognition model of instituting genuine normative force. The second is to expand the metaphysical theory of conceptual realism, already nascent in MIE, into a more robust account that also draws from *The Phenomenology of Spirit*. In tandem with these elements, Brandom deepens his processual understanding of semantic content by the Hegelian concepts of experience (*Erfahrung*), recollection (*Erinnerung*), alienation (*Entfremdung*), and forgiveness (*Verzeihung*) in ST. The final move of his impressive argumentative machinery on which I shall be especially focused on is what I shall call the 'Always Already' argument, which strives to show that the acceptance of genuine normative force is a necessary condition of possibility for being a discursive being, and hence that we are all always already practically committed to taking some discursive norms to be genuine.

My critical response to the Always Already argument and the background enabling it follows the broadly Turnerian path of argumentation that seeks to eliminate or at the very least deflate genuine normativity from (ultimately all) theoretical consideration on methodological grounds. This is fitting because my impression is that, although Brandom nowhere in ST singles him out by name, the point of the Always Already argument is directed precisely against the kind of methodological scepticism that the reductive naturalist like Turner espouses. The main idea is that scepticism is, in a sense, not self-contradictory, but self-defeating, in that in order to question the existence of genuine norms, the sceptic must 'always already' presume that norms are in force to express her doubt. Already here, we can see a certain

bifurcation of the Always Already argument into two versions, 'weaker' and 'stronger'. The weaker, 'deontic' reading claims that, given that the sceptic is *de facto* a discursive being, doubting that there are genuine discursive norms is something she ought not to do. The stronger, 'alethic' reading claims that, given that the sceptic is *de facto* a discursive being, doubting that there are genuine discursive norms is something she is not really capable of doing at all. (The possibility that Brandom, in fact, intends to assert both readings will be examined during the course of the chapter.)

Finally, a brief nondisclosure statement about my selective reading of ST. The early reception, to which the present work also belongs, has thus far centred on the exegetical-historical aspects of the book, the urgent question being how useful or faithful of a reading Brandom delivers about Hegel. This is naturally to be expected and reminds me a great deal of WRPL's early reception that was dominated by criticism of Kripke's rendering of Wittgenstein's views.⁷⁶ Yet, the more further one goes into the debate, the arguments of the phantom philosopher Kripkenstein increasingly distinguish themselves as worthy of a *sui generis* literature unmoored from Wittgenstein's exegesis. It might be obvious to all, but I predict that something very similar will happen to *A Spirit of Trust*, where the philosopher 'Bregel'⁷⁷ will come to pronounced, independent relevance once the debate learns to unweave the exegetical-historical lines of inquiry from the argumentative ones.

Unlike the case of Kripkenstein, I think that predominantly, the views of Bregel are those of Brandom as well. The nondisclosure clause that I reserve for myself is that since it is clear that ST is much more than another book about Hegel; in order to read it, it is possible – perhaps necessary – to ignore the question of how much exactly it is (really) about the historical Hegel. In the simplest practical terms, this means that I abbreviate the expositional parts starting with 'According to Brandom's reading of Hegel' to 'According to Brandom' unless an exception is for some reason required – in which case, I shall make a specific note about it. The same goes for the more elaborate historical construals, such as 'According to Brandom's reading of

⁷⁶ See in particular, Baker and Hacker (1984), Peacocke (1984), and Anscombe (1985).

⁷⁷ The portmanteau 'Bregel' comes from Pirmin Stekeler; see (Brandom 2021, 757)

Hegel's reading of Kant's reading of Rousseau'. For better or worse, it is all about Brandom.

Saying that raises the obvious demand to connect what Brandom says in *ST* with his other projects in the philosophy of language, normativity, mind, and most importantly, MIE. The positive feature of my choice to delineate Hegel out of the picture is the opening to understand how these two books interact with each other, the beginning of which I make in Section 15.3.

Another important contrast to *ST* is, of course, Kripkenstein's sceptical challenge, a connection that Brandom himself brings up in *ST*, if only in passing at the end. The surreptitious way in which the link is forged might be deceptively superficial, but I think that Brandom's theory (and in particular, the *Always Already* argument) is in deep dialogue with the sceptical challenge, the purport of the former being to solve the latter. Here, my ambitions will become more than critical, for I think that despite the shortcomings of the *Always Already* argument, Brandom's reading of Hegel delivers important insights into the post-Kripkenstein metasemantic theory.

14 INSTITUTION OF NORMS BY RECIPROCAL RECOGNITION

The purpose of this section is to describe how genuine discursive norms may be said to be instituted by reciprocal recognition. In Section 12, I already explained what the key criteria of adequacy are for such an account. The crucial issue, we saw, is to show how the process or act of institution may *fail*. For that is the only way to understand the institution as a *socio-historical* phenomenon and not a metaphysically special category of action. Analogous to how Brandom aims to explain the concept of representational success in terms of representational purport, he must give an account on the side of normative force that respects the possibility that the very institution of norms may fail to produce genuine norms. Only after that condition has been satisfied is it possible to explain how the institution may succeed.

To be explicit about this seminal point, the idea is not merely that to be successful in the ordinary sense, institutive actions or processes must be capable of failure, but more importantly, their failure must be *intelligible* prior to the intelligibility of their success. To follow the analogy to the semantic explanation of truth in terms of error (incompatibility relations), Brandom must literally explain how institutive action can fail *before* defining what it would be for the action to succeed. The simple reason for this condition is that if the institutive action is supposed to explain how norms as standards of assessment come into force to begin with, then clearly enough, the institutive action cannot be explained by a reference to the normative standards of success that would be in force already. The failure of which the instituting action must be capable is thus aptly called 'original error' since it must be a kind of failure, which precedes a concept of its own success. In other words, it must be the case that we acquire the concept of (practical) error prior to the concept of success (or correctness). Fulfilling this key condition comes into focus in the context of Brandom's discussion of Hegel's concept of experience in Section 15.

In Chapter II, I expounded tentatively on the criteria of adequacy that Brandom's brand of the institution claim is tasked to meet, but I also said something about how he aims to meet them in ST. One very general trait of the synthesising strategy is to treat both the normative statuses of authority (Hegel's 'independence'; see ST, 267) and responsibility (Hegel's 'dependence', *ibid.*) as explanatory equals, and similarly with the normative attitudes of acknowledgement and attribution. This change concerns the relations of the attitudes and statuses, but there is also something fundamentally different, at least in the very make-up of attitudes in ST as compared to MIE. The key term here is 'self-reflective', which means that the attitudes as depicted in ST essentially include a relation not only to other attitudes and norms (statuses) but also 'meta-views' regarding how the attitudes and statuses are related. The self-reflective attitudes include meta-attitudes evaluating the relation between norms and attitudes. In particular, the third part of ST is structured as a dialectical exchange between the two basic, implicit meta-attitudes, named '*Sittlichkeit*' (ethical life) and 'modernity' respectively. These issues will become prevalent in Section 16 in the context of Brandom's reading of Hegelian recollection (*Erinnerung*).

By now, an astute reader will have noticed that my proceeding in this manner is partially reverse to Brandom's order of exposition in ST, where Part One's discussion of experience comes before Part Two's discussion of mutual recognition. There are two reasons for this. The first is my overall focus on the pragmatic side of Brandom's theory, which justifies dedicating relatively little space for his discussion of semantics and the representational relation in Part One. The second is that I think Brandom's overall argument is usefully presented in two stages. Stage one amounts to a kind of a formal solution to the problem of reason's sovereignty, namely how to award an equal explanatory status to the concepts of attribution and acknowledgement (on the side of attitudes) and the concepts of authority and responsibility (on the side of statuses). Stage two is meant to go beyond the formal solution by answering the foundational question, which no formal model can solve: where is the original source of discursive normativity?

Although what I take to be my strongest critical arguments come into play only in Section 16, what precedes is not merely preparatory narrative. As discussed in Section 10, one basic problem with normative attitudes in MIE was that they were supposed to emerge from non-normative nature without presupposing abilities

describable in genuinely intentional vocabulary, which seemed to fail. In ST, Brandom takes on the challenge by providing a schematic explanation of how 'incipiently normative relations' can emerge from the non-normative base abilities – most importantly, animal desire. The desire in question is in turn supposed to be explainable in terms of practical intentionality, which (at least in my interpretation) is sufficiently describable purely as a causal-dispositional form of behaviour. I shall claim that, independently of the Always Already argument, Brandom's genealogical account suffers from the same difficulties as any other origin story of normativity, which Turner has so emphasised.

14.1 Iron, Wolves, and the Problem of Induction

In principle, how does a block of iron that displays the reliable differential responsive disposition (RDRD) of rusting in wet environments differ from a wolf responding to the availability of a lamb by eating it up? The key difference has to do with how we are to describe *to what* the thing displaying the disposition responds reliably and differently. Importantly, there are two kinds of descriptions available here: descriptions of the circumstances in which the differential response is (or has been) *in fact* triggered and descriptions of the circumstances in which it *would* (always) be triggered. For example, suppose that we have empirically established that iron rusts in wet conditions. However, all the observed conditions have also included sunlight, air, and if nothing else, then the presence of observers themselves or their instruments. How is the description of the unique conditions in which iron *would always* (i.e. 'necessarily') rust, to be individuated among the possible, rusting-irrelevant alternatives?

At stake here is, of course, the classical problem of induction and, relatedly, causation. What is relevant for Brandom is that with the wolf, the problem presents itself in an interestingly novel way. Whereas there is (usually) no question whether iron's behaviour in response to wet conditions is explainable in purely causal terms, in the case of the wolf, we are naturally drawn to apply intentional vocabulary to solve the ambiguity of conditional descriptions. Iron's behaviour is not defined by any purpose or teleological function, but the wolf's *aim* in eating the lamb is to satisfy its hunger. What distinguishes the wolf from the block of iron in this perspective is

that not all the conditions in which the wolf actually responds to the availability of an object by eating it overlap with the description of circumstances in which its hunger is satisfied. The wolf might, for example, eat a lamb that is sick, or a clever counterfeit made of plastic. No such distinction arises for rust; there is no sense in which its reliable, differential response of rusting in wet environments would fail to fulfil some purpose or desire, simply because it has no purposes or desires of its own.⁷⁸

The example of the wolf and the lamb showcases a triadic structure that Brandom calls 'orectic awareness'. The parts of the structure are:

- i) Attitude (e.g. desire; hunger)
- ii) Significance (e.g. food; a lamb)
- iii) Activity (e.g. eating in response to hunger)

The parts are related as follows. The Attitude *motivates* the Activity in the sense of (more or less) reliably activating the Activity appropriate for the Attitude. The 'appropriateness' of the Activity is determined by the *objective significance* practically attributed to the object, i.e. whether it indeed satisfies the Attitude that motivated the Activity. (*Subjective* significance means the status that is attributed to the object by the Activity that involves it regardless of whether it is successful or not.) An Attitude is 'satisfied' when it stops motivating the Activity (ST, 248).

Perhaps the most important point that Brandom wants to make by examining the very basic structure of orectic awareness as a form of primal intentionality is the notion of *error* it implicitly involves. As already noted, a block of iron (or any other merely *physical* object for that matter) cannot *fail* to rust in any sense simply because it does not aim to rust. Supposing one accepts that there are some law-like conditions that determine the unique circumstances in which iron would always rust, the task of descriptively individuating these is solely epistemic on our part. But there is a

⁷⁸ The wolf's case is, in fact, more complicated than this. At the very least, we should further distinguish between the *proximal* and *distal* descriptions of aim, purpose, or desire that motivates the wolf to eat the lamb, following the discussion of Ruth Millikan (1990). For present purposes, we can ignore many subtleties relating to the actual description and explanation of the wolf's behaviour, however.

sense in which the wolf itself can err in its practical attribution of subjective significance to a given object in that the object does not, in fact, have the objective significance appropriate for satisfying the attitude that motivated the activity. Eating a plastic lamb is a case in point.⁷⁹

The ability to register errors is a vital aspect of *learning* to adapt to an environment. The important, very general point that should now be recognised is that what is being learned, the content of learning, e.g. the disposition to avoid eating plastic lambs, is ultimately determined by whatever helps the animal to adapt to its environment. Nature is deeply opportunistic, in two senses: (a) learning is purely results-oriented (adaptation/survival); and (b) it always chooses the path of least resistance, or whatever works the best in a given context. In contrast to these somewhat anecdotal remarks stands an equally encompassing negative claim: there is no *genuine normativity* involved in explaining the learning process in either phylogenetic or ontogenetic levels.⁸⁰ In principle, the explanation of behaviour can manage with the same austere causal-dispositional vocabulary as it does in the case of the block of iron. What the intentional vocabulary of aims, purposes, and desires captures is the *particular history* of the species or the individual, which, since both are fundamentally unique phenomena, cannot be understood on the model of necessary and sufficient natural laws as with iron's rusting (Millikan 1999). The error involved in the wolf's behaviour towards the plastic lamb is hence non-normative.⁸¹ The contrasting concept of *normative error* will be discussed in the next subsection.

⁷⁹ Note that the wolf may also (at least objectively) err in registering the cause of dissatisfaction. Many animals are disposed to associate sickness they later endure with certain foods they consumed earlier, as opposed to the place, time, or some other environmental variable, and thus learn to avoid the type of objects henceforth, even when it was the environment that caused the sickness. See Millikan's work (1990) for more discussion.

⁸⁰ This is not the place to go into the deep, difficult question of whether *any* non-linguistic animal behaviour can be described in *any* genuinely normative terms, e.g. as displaying regret. Nor can I touch upon examples of learned behaviour, such as the ones displayed by the celebrated New Caledonian crow, which appears to exemplify a form of *culture* in the sense of instrumentally designing tools for specific purposes (Bayern, A.M.P.v., Danel, S., Auersperg, A.M.I. *et al.* 2018). These and other numerous examples would require a much more fine-grained discussion than what can be mustered here, though my assumption is that that discussion requires getting clear about the more general issues about the relations of normativity and causal-dispositional explanation that are my (and Brandom's) focus.

⁸¹ The wolf is still notably different from iron in another respect aside from having a unique history. Even if one does accept that there are natural laws governing the behaviour of iron (and by extension, of the wolf considered purely as a physical object), it does not follow that the task of descriptively

To sum it up, how does the classical problem of induction change when applied to the wolf as opposed to the block of iron? First, in order to explain the wolf's behaviour, but not iron's, we must refer to its *history* (phylogenetic and ontogenetic, depending on whether we are talking about types or tokens). Second, part of the history must be understood as a prolonged process of *learning* from failures to adapt to the environment.⁸² Third, from these two points, it follows that the wolf, unlike the block of iron, can *change* its future dispositional behaviour in ways that are not necessarily predictable by any nomological model.

However – and this is crucial – the *methods* for explaining the wolf's behaviour do not *in general* have to differ from the causal-dispositional concepts available to explain the behaviour of iron, supposing that we understand the 'causal-dispositional' methods broadly enough. The 'causal-dispositional' explanation I am appealing to here is obviously left mostly a placeholder; what matters is the contrast to the genuinely normative explanation, which Brandom thinks can (and must) be distinguished from it in the case of discursive, linguistic behaviour.

14.2 Recognitive Hierarchy

Having compared the behaviours of the iron block and the wolf, it is time to turn to more complex cases. Brandom claims we can understand genuinely normative behaviour as something *arising* from the more basic kind of orrectic awareness described above. If we understand the behaviour of the wolf towards the lamb as an instance of 'K-taking' (taking something *as* something per the triadic structure of orrectic awareness), the hierarchy of behaviour that Brandom claims arises from the following:

individuating the content of the wolf's learning is simply epistemic as it is with iron, or that the right answer would 'already be there' at the metaphysical level. For it is not obvious that the wolf's disposition to learn is everywhere determined in a law-like manner; it is not determined what the wolf (much less any particular wolf with its particular history) will learn after eating the plastic lamb. For all I know about him, the wolf might even (learn to) like the taste.

82 We should, of course, avoid hastily concluding that every feature of the wolf, be it behavioural, physiological, or anatomical, would necessarily have to hold some purpose that either has helped it adapt in the past or currently helps it to adapt. In practice, it is often daunting to discover just which features have been selected for, to what end, and which are simply byproducts of (de)selection.

- (i) K-taking (triadic structure of orectic awareness)
- (ii) Taking something as a K-taker (specific recognition)
- (iii) Taking something as a taker of K-takers (general recognition)

Moreover, Brandom claims that we can understand each subsequent stage in this hierarchy, according to the triadic structure of Attitude, Significance, and Activity. How this happens is, first, by inserting the behaviour (Activity) at the previous stage (e.g. K-taking) into the Significance slot of the subsequent stage. Two questions can then be formulated. First, what is it that one must do in order to treat something as a K-taker; and after that, what is it that one must do in order to treat something as a taker of K-takers?⁸³ Second, we must answer a corresponding question regarding the Attitude slot for each stage, so that we can state what (a) motivates and (b) determines the success of the Activity of treating something as a K-taker or as a taker of K-takers. The second-order takings at stages (ii) and (iii) are what Brandom calls 'recognitions' (ST, 249).

Before proceeding any further, let us examine the general purpose of the account under consideration. Brandom's idea is that once we have answered the two questions above, it becomes possible to understand Hegel's claim that 'Self-consciousness is desire' as follows: (a) mutual (or reciprocal) recognition is necessary and sufficient for the dyad of recogniser and recognised to count as self-conscious beings, and (b) the Attitude slot for (iii) is filled by the 'desire for recognition'.

The point on which I shall focus the most here is that as a part of this major exegetical argument, Brandom claims that reciprocal recognition is a genuinely normative relation that arises from non-normative orectic awareness. I am inclined to interpret this as a kind of a genealogical claim: orectic awareness is a historical stage of self-conscious beings, with specific and general recognition being the necessary waypoints to self-consciousness via mutual recognition. (They are *necessary*

⁸³ Another, more streamlined term for 'taking something as a taker of K-takers' is to take something simply as a taker, or as a *subject* for which things can be *for*, which is *conscious* of things as having some subjective significance. I have strived to avoid this Hegelian terminology here because licensing myself to it would require a lengthy separate discussion. I think my main points can be made independently of the exegetical questions, as I already stated in the Introduction to the chapter.

in the retrospective sense that without these waypoints, self-conscious beings could not have arisen, not that it is necessary for there to be self-conscious beings.) However, I shall argue that there is a piece missing from the story as told by Brandom, at least in the sections where the argument first appears.

14.3 Argument against Authority

Let us start by observing what, for Brandom, fits into the Significance slot in the triadic structure examined above with the second-order dispositional behaviour called specific recognition, or taking something to be a K-taker. The answer is *authority*:

This is the decisive point. My taking *your* K-response to have been authorized by a K-desire that serves as a standard for the success of your K-taking, and taking that K-response to have been *correct* or *successful* by that standard is my acknowledging the *authority* of your K-taking, in the practical sense of being disposed *myself* to take *as* K the thing *you* took to be a K. (ST, 252)

The problematic thought that Brandom here espouses is the application of the normative term 'authority' to describe the second-order disposition of subject A to treat B's disposition to classify an object as K, in the sense that A comes to have a similar disposition to treat the (type of) object as K. The fact that Brandom thinks underway here is a qualitative shift in takings (from non-normative to normative) becomes apparent one page later:

Looking back at the most primitive sort of preconceptual recognition of others from the vantage point of the fully developed, conceptually articulated kind brings into relief the crucial boundary that is being crossed: between the merely *natural* and the incipiently *normative*. In the merely orrectically aware animal, *desire* is a state that *motivates* and *regulates* responsive activity *immediately*. [...] The recognizer accordingly takes up a more distanced, mediated, abstract attitude toward these significance-generating attitudes. The recognized creature's attitudes are treated in practice as *assessing* the *correctness* of practical responsive classifications, as *licensing* or *authorizing* the responsive activity—in the first instance in the case of the one recognized, but then also on the part of the recognizer, who merely attributes the attitude to the other. The relation between the attitude the recognizer attributes and the activity he himself engages in is a *normative* one. Even in the most primitive cases it is intelligible as the acknowledgement of *authority* rather than mere acquiescence in an impulse. In treating the attitudes of the recognized other as having authority for those who do not feel them, the recognizer implicitly accords them a significance beyond that of mere

desires: as *normatively* and not merely *immediately* significant attitudes. (ST, 253, footnote omitted)

It is difficult to discern Brandom's exact justifications for the shift from impulse to authority, or from nature to normativity, in these sections (ST, Ch. 9, Sec. I–II). The major pivot appears to be that the recognising subject A does not herself *feel* the desire, which she attributes to B as motivating his behaviour, and thus accordingly has a more 'distant, mediated, abstract' attitude towards B's desire considered as a standard for evaluating the success of behaviour (both A's and B's). Since B's desire (as practically attributed by A) cannot *immediately* motivate A's behaviour, its role regarding A's behaviour is understood as 'incipiently normative'. The seminal question, however, is: why should this be counted as *normative* evaluation even in an 'incipient' sense? For all that seems to go on between A and B is that A practically treats B's behaviour towards objects (of a given perceived type) as a *reliable indicator* that they are Ks. This does not seem to notably differ from B's disposition to treat objects as Ks. In particular, it seems that we can adequately account for the *errors* of practical attribution of Significance in the same way for both first-order and second-order dispositions.

To illustrate, recall what it means for K-taking to go wrong, i.e. to register an error. The wolf perceives an object that appears to it as a lamb, which is classified as food by the wolf. As the wolf is hungry, he thus proceeds to eat the lamb. However, it turns out the lamb was made of plastic, which fails to satisfy the wolf's hunger, leading to a registration of error. What it means for error to be registered is that the wolf's hunger is not satisfied, which means that the wolf proceeds its Activity loop of finding new sources of food that it does not associate with plastic, for (if all has gone well) the wolf has learned not to eat plastic lambs.

Now, consider how the same picture can be used to describe the second-order case. A is practically disposed to treat B as a K-taker. A perceives B treating an object as K. A thus gains a disposition to treat the (type of) object as K herself. Since A happens to have an Attitude that motivates the Activity towards K-taking, she engages in K-taking, possibly by fighting B for the same object or not, if the sources are plenty. However, it turns out that the objects, which A took B to have classified as Ks, are not in fact Ks, which means they do not satisfy the Attitude that motivated A's K-taking, meaning that A registers an error and (if all goes well), learns not to

take the objects as Ks henceforth. However, what A may *also* learn is not to treat B as a reliable indicator of (certain types of) objects as Ks. In Brandom's terms, B loses his 'authority' as a K-taker for A.

Now we must ask: where is 'normativity' in this picture? Why is the term 'authority' more than a metaphor for B's Attitude, as attributed by A to causally regulate A's behaviour? In both the first-order and the second-order cases, the error is non-normative, which means they are explainable purely as a change of behavioural dispositions due to certain experiences. The important point is that in the second-order case described above, there is no situation where A treats B as *incorrectly* treating an object as K. The reason for this is that what it means for A to treat B as a K-taker just is for A to treat B's Activity as a reliable indicator of Ks, which means that A's K-taking is causally regulated by B's behaviour. If A does not treat B as reliable in this way – perhaps because she has learned to know better – A cannot register B's activities as K-taking in general, and hence cannot register his *incorrect* K-takings in particular. But if there are no situations where A would register B as engaging in K-taking *incorrectly*, it seems difficult to say why this relation should be counted as normative, in the sense that A registers B as doing something he *ought not* to do, even in just an instrumental sense of 'ought'. But that is a necessary condition for something to count as a *normative error*, as was discussed in Section 12.

But perhaps this is too quick. The decisive question that needs to be resolved is *how* A comes to treat B's behaviour as K-taking to begin with. Two interpretations seem to be available. On the first interpretation, which is implicit in my alternative treatment of the example above, A's method for determining B's behaviour as K-taking is purely *results-oriented*: in the past, it has happened that B's behaviour, *however it is independently described*, has (more often than not) led to the discovery on A's part that certain objects are Ks. On the second interpretation, A registers B's behaviour as K-taking simply because it is (registered as) the same behaviour A herself would engage in as K-taking, *irrespective of what A projects as the success of the behaviour in a given context* either for herself or for B. Summed up roughly: does A register B's behaviour as K-taking by relying on the *output* or *input* conditions?

To be explicit, this question concerns the *schematic order* of the *procedure* by which A is to recognise B as a K-taker (i.e. specific recognition). To go over the question in

detail, let us first make explicit the elements of the process, which Brandom thinks are necessary and sufficient to explain specific recognition (i.e. taking something as a K-taker):

- (1) One must attribute an activity that one takes to be what it is for the other to be responding to something *as* a K.
- (2) One must attribute a desire or other attitude that one takes to *licence* or *authorise* responding to things *as* Ks—that is, by engaging in that activity.
- (3) One must acknowledge in practice a distinction between *correct* and *incorrect* responses of that sort, assessed according to the attributed attitude that authorises responses of that kind. (ST, 251)

Brandom notes that it is a necessary condition for the specific recognition of something as a K-taker that the recogniser herself is a K-taker: 'That is, my taking you to be able to treat things as *food* is my taking it that you respond to some things with the same behaviour, *eating*, with which I respond to food' (ibid.). However, even if it necessarily takes one (K-taker) to know one, I think it is clear that being a K-taker is not generally a *sufficient condition* for the ability to recognise someone as a K-taker. (To be clear, Brandom does not claim that it would be.)

Brandom does not say this explicitly, but I think he opts for the 'input' strategy in explaining what it is for A to register B's behaviour as K-taking, e.g. eating. What this means is that A can 'match' the behaviour B exhibits as of the same type as what A exhibits towards objects, which A classifies as food. As a consequence of this matching, A then attributes the Attitude, hunger, that is objectively appropriate for the Activity, and then projects the success (or correctness) of the Activity in the context of the objects that B happens to be eating. Alternatively, without the ability to project (e.g. because A does not have data to register whether the objects B eats are edible), if all goes well, A learns from B's behaviour whether the objects are edible.

Note that the 'matching' of behaviour (my term, not Brandom's) on A's part must be explainable in non-functional terms, i.e. without reference to the attributed Attitude that serves as a standard of success for the Activity, for the Activity is supposed to be the first procedural element in how the Attitude is individuated. Moreover, the matching must also be explainable without reference to A's practical

classification of the particular objects B happens to be eating as Ks, for it is otherwise impossible for A to register B as *incorrectly* responding to an object as a K. If A's matching of B's behaviour with what A considers as, such as eating, depends on whether A would already be disposed to eat whatever B happens to eat, then if A registers B 'eating' something that A registers as inedible, A is disposed to de-match B's behaviour with her own eating rather than take B as eating incorrect things.

We can now enter the comparison of the 'input' and 'output' explanations of how A comes to specifically recognise B as a K-taker. The input interpretation appears to explain how A can, after all, register a sort of incipient normative mistake in B's behaviour. Perhaps, after the example above, A learns to avoid classifying the objects as Ks, whereas B, being a slower learner, does not. So, it becomes possible for A to continue recognising B as a K-taker, only a *bad* one, which is a more sophisticated category than what was available to A before. For before, A could either take B as a K-taker or not, where 'K-taker' is defined along some parameters of reliability. By becoming aware of B's mistaken classification, A, by that token, comes to be aware (i.e. register or recognise) an *instrumental* form of normativity, which is plausibly what Brandom has in mind with 'incipient' normativity.⁸⁴

Moreover, Brandom thinks that the *only* way A can *learn* from B's K-taking is if A attributes to B's Attitude the *authority* to regulate A's K-taking, to classify objects as food or not-food based on whether eating it satisfies B's hunger in the sense of leading to the cessation of eating. A effectively treats B as an authority on K-taking in general, as making a general practical *judgement* about what is edible for everyone, or perhaps to everyone indexed to a certain similarity class (e.g. an authority for other wolves, or an authority as to what is an edible mushroom). And because attributing potential *incorrect* attributions of subjective significance by B is a necessary condition for A to treat B as a K-taker, it follows that 'incipient normativity' is a necessary element in explaining the second-order dispositional behaviour, or the shift from impulse to authority.

⁸⁴ The normativity in question need not be instrumental but could be more 'deontic', as it were. Basically, the question is whether A treats B's attitude as laying a judgement about an object being 'good for eating' or (something like) 'right/correct to be eaten'. Here, nothing of great relevance hinges on how exactly we interpret 'incipient normativity' as normativity since the main question is whether we need normativity at all.

The main point that I have sought to highlight with my alternative explanation of A's ability to specifically recognise B as a K-taker by relying on the output conditions is that, in order to learn from B's K-taking, it is *not* necessary for A to practically attribute an Attitude that motivates and regulates B's K-behaviour. It follows that it is neither necessary for A to be able to register B as having made a *mistake* in his practical attributions of subjective significance in order to learn from his K-behaviour, so that a 'genuinely normative mistake' is not required for an explanation of the second-order case. In this perspective, what ultimately matters for A's learning is that the objects she comes to eat really are food, not whether B is *correct* or *incorrect* (is/has an authority) in his attributions of subjective significance; and for this, it is not necessary for A to be able to project incorrect classifications on B's K-takings.

It is a consequence of the output reading that A's ability to recognise B's behaviour as K-taking is in principle independent of the subjective significance, which B himself actually attributes to an object. Because B's behaviour in this interpretation counts as K-taking by A only in virtue of the fact that it reliably leads A to successfully treat certain objects as Ks, it is perfectly possible for A to be generally successful in her second-order attributions, although B is not, in fact, engaging in (the same type of) K-taking at all, i.e. does not attribute to the objects the subjective significance as (the same) Ks. For example, suppose that the behaviour, which A attributes to B, with the subjective significance of eating for B has the subjective significance of finding good nest-building material. If it happens to be true that the objects, which, for B, make for good nest-building material, double up as nourishing food for A, then A's treatment of B as a reliable food-indicator is successful regardless of the fact that A and B are indeed motivated by very different Attitudes and engage in very different Activities.

But is the second-order behaviour thus described appropriately called 'specific recognition' anymore? I take it that, for Brandom, it would not be simply because, although A is in a sense relating to B as an 'authority' on food sources, the resulting success is purely accidental and, perhaps even more revealingly, *asymmetrical* in the sense that how A relates to B's behaviour is very different from how B relates to his own behaviour. This reading is, I think, present in the following passage:

What it is for it to be K-takings (and not some other significance or no significance at all that you are practically attributing to things by responding to them in that way)

that I *take* your responses to be consists in the fact that it is *my* K-taking responsiveness (and not some other activity) that I am conditionally disposed to extend to the kind of objects that satisfied your desire. The link by which the specifically recognized one's activity is assimilated to that of the recognizer is forged by the *interpersonal* character of the specific authority of the recognized one's successful takings, whose acknowledgement is what specific recognition consists in. The *only* way the recognizer's orectic classifications can be practically mapped onto those of the other so as to be intelligible as implicitly attributing specific desires, significances, and mediating responsive activities exhibiting the TSOA [triadic structure of orectic awareness], is if the *authority* of the assessments of responsive significance-attribution on the part of the one recognized is acknowledged in practice by the recognizer. (ST, 251–252)

In other words, according to Brandom, A cannot properly be said to (specifically) recognise B as a K-taker unless A takes it that what motivates B's Activity is the *same* K-Attitude as what motivates A's K-taking. The important question, of course, is not whether this is a good definition (or an exegetical reading) of 'specific recognition', but whether it is required in order to explain the kind of second-order dispositional behaviour under discussion. I have argued, based on the output reading, that in order for A to treat B as a reliable indicator of objects as Ks, it is not necessary for A to attribute to B *any* Attitude, much less the *same* one as what motivates A's K-taking. *The crucial question then becomes: can the output interpretation be taken as a competing account of what it is for A to take B's behaviour as exhibiting the triadic structure of orectic awareness (TSOA)?* If the answer is yes, then Brandom is wrong when he, in the quote above, claims that the *only* way for A to specifically recognise B as exhibiting TSOA is by practically acknowledging the *authority* of his K-attitude. For as we just saw, the output reading does not imply that A must attribute to B *any* Attitude, much less *the same* K-attitude she is motivated by, in order to practically treat B as a reliable indicator of objects as Ks, thus a K-taker in the output sense.

There are two possible objections, at this point, to my argument that must be considered. The first is the transcendentalist insistence that 'authority' just is a conceptually necessary ingredient in the description of what specific recognition is, and since authority is a normative concept, so must specific recognition be. My alternative account does not rival Brandom's so much as speak past him. But as I already noted, this is not (should not!) be the pivot on which the argument turns, for in that case, the objection must bear all the general burdens of transcendental

arguments discussed at the end of the previous chapter. That is why I shall ignore this first objection here.

The second objection states that what Brandom has in mind in claiming that the normative term 'authority' is a necessary ingredient in describing the dynamics of specific recognition is simply that there is no other way to explain how A *can* attribute specific K-attitudes to B. This objection effectively admits that my alternative (sketch of a) 'reliable indicator' account may plausibly work as an explanation of what it is for A to *learn* of B's behaviour without presuming that A attributes the same or any K-attitudes to B. Yet, just because it can be shown that attributing K-attitudes is not necessary for learning, it remains to be shown how, in my view, it is possible for A to attribute to B specific K-attitudes, which *prima facie* is something to be explained.

And here, I think the issues multiply and become too complicated to be satisfactorily settled in this context. Admittedly, I do not have a full working story of how the practical attributions of specific K-attitudes might work without relying on normative terms, such as authority. But from that alone, it does not follow that there is no such an account available. And of course, it would have to be specified what kind of behaviour exactly should count as a pre-theoretical measure of success for the hypothetical account, in what context, etc.

Moreover, there are two further critical remarks I want to make about Brandom's claim that:

The *only* way the recogniser's orctic classifications can be practically mapped onto those of the other so as to be intelligible as implicitly attributing specific desires, significances, and mediating responsive activities exhibiting the TSOA, is if the *authority* of the assessments of responsive significance-attribution on the part of the one recognised is acknowledged in practice by the recogniser. (ST, 252)

First of all, at least in the immediate context, he does not actually provide any reasons supporting the claim that the authority account is uniquely fit to explain the relevant second-order dispositional behaviour. This is what one should expect, presuming that Brandom thereby denies that there is available any non-normative alternative account to explain how A can practically attribute specific K-attitudes to B. Second, what is telling in the lack of explicit reasons is the term 'intelligibility' on which the denial turns. Now, 'intelligibility' is often used as a transcendentalist code word for

'conceptually necessary', which leads to the usual problems. But if Brandom is not here committing himself to the transcendentalist defence of the necessity claim, his denial that there could not be an alternative non-normative explanation for specific recognition is without the force of reason.

Before moving on, there is an important addition to be made to my argument against authority. It may well happen that sometimes it is in A's interest not merely to *passively* find subjects, which function as reliable indicators of desirable Ks for her, but to *actively* affect reliable K-taking behaviour in subjects that she is already familiar with. This is the key observation made recently by Rémi Tison (2022) in his 'active inference' ecological account of social and linguistic normativity that draws from MIE. Briefly, Tison argues that we explain the emergence (or 'institution') of social and linguistic normativity by the observation that species often adapt to their environment, not by changing themselves, but by changing their environment somehow. Since the important ecological niche of humans consists of their social environment, it is often in their individual and group interest to make that environment more predictable; for example, by sanctioning the perceived deviant use of various proto-linguistic communication signals to minimise prediction errors about, e.g. the likely location of a mammoth. The sanctioning behaviour is driven, according to Tison, by the 'normative expectations' of a group, and he uses the terminology of commitments and entitlements to cash in the details of his just-so story about the possible emergence of linguistic normativity.

The important point here is that my proposed 'output reading' is compatible with the key idea of active inference framework, which is that A sometimes seeks to actively instil predictability into B's K-taking by various sanctions. For all that is arguably really going on in such cases is that A is disposed to increase the statistical likelihood that B's behaviour, however independently described, results in outcomes favourable to A's K-taking; a result that does not in principle demand that A attributes to B the *same* attitude as what motivates A's K-taking, much less that it would have to be *authority* which A thereby attributes. In my view, Tison's account does not adequately address Rosen's (1997) criticism discussed in Section 10.1: we cannot define normative terms, such as 'sanctionable deviance', only in terms of how the group actually reacts to *de facto* deviance from a putative norm because the group sometimes fails to sanction the putatively deviant behaviour due to various

contingent factors, such as ignorance. In order to talk about *genuine* normativity, we would have to show that the sanctioning itself is somehow authorised, which then brings us to the core of the problem of reason's sovereignty. Moreover, the deeper point that my argument in this section has sought to press home is that in order to explain the sanctioning behaviour as *observed in action*, all that is needed is the output reading of K-taking.

14.4 Robust Recognition

My critical argument above should not be understood as by itself (aiming to be) fatal to Brandom's account of TSOA and specific recognition. Rather, my main goal was to show that the problems, which Turner and others have raised against transcendentalist arguments for normativity, carry on as an active issue for the theory of ST – at least, at this stage of exposition. The apparent stalemate rehearsed above arises when the anti-normativist alternative explanation is met by the normativist with the allegation that it is incomplete at least in the empirical sense of not accounting for all the effects, which the normativist wants accounted for (to wit, how is it possible to attribute specific K-attitudes to others?). But this is only an appearance, for the more pressing question presented by the anti-normativist is why we should think there is no such alternative account available. After all, in a sense, there *must* be such an account somewhere unless we are to think that specific recognition is not underlined by any causal-dispositional mechanism that is by itself sufficient to explain all the observable effects we want explained. Indeed, Brandom does not deny that such accounts, in principle, exist for any human behaviour, as we shall see in Section 16. His ultimate reasons for insisting that the causal-dispositional account misses something important have not yet been discussed, for I think they essentially involve the reading about *Erinnerung*, towards which we are moving.

Before we arrive to that point, in this section, I will discuss another related problem concerning 'robust recognition', which is supposed to be the final ladder in the recognitive hierarchy. This is also where, as we recall, Brandom aims to synthesise the preceding models of institution, named as 'the Kantian Autonomy model' and 'the Queen's Shilling model'. His general strategy for this, we have now seen, is straightforward and usefully presented in a table (of my design):

Table 1. Models of discursive sovereignty

The Model	Kantian Autonomy	The Queen's Shilling	Mutual Recognition
Privileged attitude & status	Acknowledgement & Authority	Attribution & Responsibility	Attribution & Authority

Whereas the Kantian model starts its explanation with subjects acknowledging authority over commitments, and the Queen's Shilling model starts with others attributing responsibility over commitments to the subject, the mutual recognition model starts with the subject attributing authority to another subject. Authority over what? In the first instance, the authority over K-takings, which is called specific recognition. In order to move up a step, we repeat the process in the shift from K-taking to taking something as a K-taker by inserting the whole structure of specific recognition into the Significance slot of what is now called robust recognition. Hence, what the subject A attributes authority over, at this third stage, to B is the authority to recognise C, not only as a K-taker, but also as a taker of K-takers. The ladder ends here, for it adds no further recognitive power to recognise someone as a recogniser of other recognisers of recognisers (ST, 256).

If we start from the attitude of attributing authority to another, how can we construe the other combinations of attitudes and statuses on this basis, namely the acknowledgements of authority or the responsibility and attributions of responsibility? To exemplify, we can understand the undertakings of responsibility being entailed by the attributions of authority as follows. In specifically recognising B, A treats B's K-takings as having authority over her in the sense that she is responsible for acknowledging B's K-takings, in practice, e.g. by eating what B eats. So, the second-order recognition of authority entails the first-order responsibilities towards whatever occupies the Significance slot at the first-order level. Similarly, in robustly recognising B, A commits herself (becomes responsible) to recognise anyone whom B recognises as a K-taker or as a taker of K-takers. Again, the higher-level attributions of authority entail lower-level undertakings of responsibilities. And since acknowledging a responsibility just is to make explicit the undertaking, we have

essentially construed the combination of the acknowledgement of responsibility from that of the attribution of authority⁸⁵ (ST, 255).

To say that the combination of attitudes and statuses as the *attribution of authority* is 'privileged' means that it cannot be explained by reference to other combinations of attitudes and statuses. For example, as we saw in Section 12, in the Queen's Shilling model, the subject's authority to acknowledge commitments is a derivative notion explained by the attributions of responsibility by others, namely the responsibility to justify the acknowledged commitment if appropriately challenged to.⁸⁶ The question that remains for the Queen's Shilling model is how the community members can have the *authority* to attribute responsibility to each other. Brandom's answer, examined in detail in the last chapter, is to explain this by a further attribution of authority from outside the community, with all further levels later collapsing into a single community of rational-logical subjects: a move that has been criticised as inadequate, as we saw in Chapter II.

It is now noteworthy to see that the model of mutual recognition faces exactly the same question regarding the explanation of the privileged combination of status and attitude. Following Georg W. Bertram (2020, 77), we can ask how subjects can come to have the authority to (robustly) recognise others as having authority to begin with. Since every genuine normative act must be open to evaluation as to its correctness, any attempt to recognise someone as authoritative must itself enjoy some authority in order to be a genuine normative act. Brandom is right to say that *if* robust recognition is *de jure* transitive and *de facto* symmetrical, then it is, by that token, also self-reflexive *and* authoritative all the way down since my authority to recognise you as authoritative is made authoritative by your recognition of me as authoritative (ST,

⁸⁵ In this work, I have not touched on Brandom's theory of logic other than mentioning that it starts from materially good inferential relations and seeks to understand formally good ones on that basis. Thus, it is not surprising to see this treatment extended to deontic logic, in which the terms 'authority' and 'responsibility' have their home, and which, from the viewpoint of classical deontic logic, behave somewhat idiosyncratically here. Since my main purpose in this work does not directly concern Brandom's logic but rather his semantics and pragmatics, I continue to leave the matter of 'construing' as it may be. In any case, the 'construal' of the acknowledgement of responsibility from the attribution of authority is not quite so linear as portrayed here, yet for my purposes, the gist of it should do.

⁸⁶ To complete the picture, Brandom also gives a schematic account of how, in the Kantian autonomy model, it is possible to construe, e.g. the attributions of responsibility based on the explanatorily privileged acknowledgement of authority (ST, 271).

254). But as Bertram observes, Brandom does not, in the immediate context (ST, Ch.8, Sec. VI), explain *how* robust recognition can be *de facto* symmetrical.⁸⁷ Moreover, there is an obvious chicken-and-egg problem involved here: if A's authority to recognise B's authority depends on B's *actual* recognition of A's authority *and vice versa*, neither's recognition of the other's authority can be authoritative *before* the other's is.

This dilemma has two principled solutions available that we have already discussed, both of which have the form of foundationalist regress-stoppers. The Kantian Autonomy model posits a metaphysical power for the subject to make norms binding via her acknowledgements, which amounts to a transcendental explanation. The Queen's Shilling model opts for reduction instead. Clearly, the model of mutual recognition (as an *Aufhebung* of the two) cannot accept either outcome. As we shall later see (in Section 16), Brandom's way out of the formal dilemma has to do with his theory of action, which involves the theory of *Erinnerung* and the temporal dynamics of institution. What has been shown thus far, then, is that to truly synthesise the Kantian and the Queen's Shilling model, it does not suffice to combine their respective views of normative attitudes and statuses, which is essentially achieved by robust recognition. What also requires synthesising are certain views regarding the *temporality* of norms and attitudes.

⁸⁷ That robust recognition is arguably transitive was seen above in conjunction with explaining how the higher-order attributions of authority entail lower-level undertakings of responsibility. Again, probing any deeper into this aspect of Brandom's theory would require a discussion of his (deontic) logic, which I have delineated out of the picture in this work.

15 ON EXPERIENCE

Back in Section 9.3, I summarised Brandom's extension of MIE's doubly perspectival account of conceptual content into a triply perspectival version, which includes a historical dimension. The central lesson, read as Hegel's improvement on Kant (or as Quine's improvement on Carnap), was that instead of a two-stage process where absolutely determinate contents are first selected for and then applied in judgements, we should understand the process of determining conceptual contents as concurrent with the process of application, where these two sides appear from opposite temporal perspectives. Retrospectively, it appears to the acting subject that her process of experience of rectifying incompatible commitments is epistemic in nature, in the sense that she tries to discover how things have already always been regardless of her semantic decisions in order to decide on a present verdict. Prospectively, once the decision has been made, it may appear either to the subject herself or others following her that she not merely found but made a novel contribution to the content of a concept by choosing to apply it one way rather than another. I also briefly discussed the conundrum of how these two perspectives are to be reconciled in the same subject, though no clear conciliation could be reached.

The purpose of this section is to examine Brandom's latest attempt to explicate his one-stage, historical process of experience. My overarching aim, however, is not to study experience (*Erfahrung*) for its own sake, for as we shall see, it ultimately concerns the issue of semantic objectivity, whereas my aims lie with what I called the pragmatic objectivity of normative force in Section 12. The reason to include an extended discussion of experience is that the *recollective* process that is embedded in experience is an important piece in Brandom's answer to the foundational question about the original authority to attribute authority to someone else (i.e. how mutual recognition can be *de facto* symmetrical), which was brought up at the end of the previous section. Keeping the pragmatic-semantic distinction in mind, it will be helpful to distinguish between the *semantic recollection of content* and *pragmatic recollection of intentions*.

In both cases, what 'recollection' (Hegel's *Erinnerung*) means is, roughly, that in order to rectify a registered incompatibility in her present commitments, the subject must present those commitments as elements in a diachronic, historical process, which culminates in the new, rectified constellation of commitments she acknowledges in the present. This idea forms the core of Brandom's *Vernunft*, one-stage incompatibility semantics. On the side of pragmatics, recollection retains the same form but changes its target from theoretical commitments (i.e. beliefs) to practical commitments (i.e. intentions). As we shall see, since, for Brandom, symmetrical recognition is something that is achieved not only in but also over time (indeed backwards in time), the key issue is how the recognising subject can (and whether she must) attribute to the recognised one an intention to follow some genuine norm in her actions; in other words, to attribute an intention to do what is right *because it is right*.

At this point, it is good to take stock of the big picture that this chapter has been painting. Brandom's strategy for explaining how genuine discursive norms can arise from discursive attitudes starts with the Hegelian mutual recognition model, which is a kind of a synthesis between the earlier Queen's Shilling and Kantian autonomy models. In particular, the synthesis takes the explanatorily privileged combination of attitudes and statuses to be that of the attribution of authority. Furthermore, the mutual recognition model includes three different levels of recognition, with the final, robust recognition standing for the formal solution to the problem of reason's sovereignty. It is merely 'formal' because at this stage, supposing that robust recognition is *de jure* transitive and *de facto* symmetrical, it is thereby genuinely normative, i.e. the subjects are really authorised to recognise each other. But we have not yet been told how the *de facto* problem is to be solved, which arguably is the philosophically substantial issue. That is where Part 3 of ST makes its contribution, for it is meant to offer a story about the pragmatic recollection of intentions, following the model of the semantic recollection of content.

The section will proceed as follows. In ST, Brandom expands on both his exegetical and argumentative justifications for attributing the shift from Kant's two-stage account to Hegel's one-stage improvement. My focus in Section 15.1 will be on the argumentative side. This reintroduces us to the idea of conceptual realism, which we recall I attributed as Brandom's implicit solution in MIE to solve the problem of

semantic objectivity, or how the world of objects itself can exercise authority (thus function as a normative referential anchor) for our conceptions of it. In Section 15.2, I will explain how Brandom further develops this part of his theory, along with a brief examination of some problems with it. Finally, in Section 15.3, I shall make explicit certain comparisons between MIE and ST on the process of experience, and how these works can indeed be seen as radically diverging on some key methodological points, especially regarding the nature of normative attitudes. At this point, the table should be set to enter the culminating debate of this chapter, namely Brandom's proposal for solving the problem of reason's sovereignty by a pragmatic recollection of intentions.

15.1 Four Semantic Criteria of Adequacy for Epistemic Theories

The crucial feature that unites the two-stage representational theories of Descartes, Kant, and Frege, according to Brandom, is that they posit a categorical difference between representings and representeds. The categorical difference is epistemic in kind: it concerns the *intelligibility* of the two sides, or how we come to cognitively grasp or understand them, respectively. The two-stage representational theories in effect claim that while representings are intelligible intrinsically, representeds are intelligible only extrinsically, in at least two senses. First, to be intrinsically intelligible means that (i) a subject can come to grasp representings without representing them by other representings, i.e. the subject's relation to representings is not itself representational, and (ii) to be intelligible or graspable is an essential part of what it is to be a representing. Second, to be extrinsically intelligible means that (i) a subject can come to grasp or understand representeds *only* by means of the representational relation, and (ii) to be grasped in such a representational relation is not an essential part of what it is to be a represented; the status of being a represented is not essential to the object being represented⁸⁸ (ST, 42–43).

⁸⁸ We should not confuse the distinction 'intrinsic/extrinsic', which are my terms, with that of 'mediate/immediate'. The two-stage representational theories as such can go either way as to whether the subject's grasp of representings is immediate or mediate: representings can be grasped immediately (i.e. atomistically) without relation to anything else or to other representings or mediately (i.e. holistically with relations to other representings) (ST, 42).

The rationale for defining what it is to be intrinsically intelligible is clear. The reason the subject's relation to representings cannot be representational everywhere is simply that it would lead to a regress of representational relations, since the secondary representational relation between the subject and her first-order representing would in turn have to be supported by a tertiary representational relation between the subject and a second-order representing, and so on. Similarly, it is intuitively clear (if not uncontested) that for anything to be a representing, it must be capable of being a representing *for* someone, i.e. graspable, while to be a represented, it is not similarly essential (barring full-blown subjective idealism) that the object is (even capable) of being represented for someone.

The first distinction has an important, immediate consequence:

In a formulation that was not extracted explicitly until centuries later [from Descartes' time] by Josiah Royce, if even *error* (*misrepresentation*), never mind knowledge, is to be possible, then there must be something about which error is *not* possible—something we know about *not* by representing it, so that error in the sense of misrepresentation is not possible. (ST, 40)

However, the relation of subjects to their representings is to be explained, one necessary condition is that representational error is not possible in that domain, not because the representational relation would as a rule succeed here, but because the relation cannot be representational. If error is possible in the relation between a subject and a representing, it must be non-representational in kind. And so – and this part is crucial – because the two-stage representational view understands the subject's epistemic relation both to representings and representeds as something consisting of *knowledge-that*, it follows that there are two crucially different kinds of propositional knowledge: knowledge of representings and knowledge of representends. The major, defining difference between these two kinds of knowledge-that is that it is impossible to be in error about knowledge-that relating to representings, for that is the privilege of knowledge-that about representeds. Even though Descartes understood that the subject's relation to her representings cannot itself be representational, he took it that it still counted as a kind of propositional knowledge. Moreover, it was in virtue of this incorrigible (error-free) knowledge of representings, which, for Descartes, made possible the knowledge-that of *res extensa*, the represented objects. The reason the knowledge of representings must be incorrigible, hence special, is that it must be free of representational errors, i.e., the

only relevant kind of error possible for propositional knowledge, according to the two-stage theories.

Brandom's first major claim in ST is that (according to Hegel) the discrimination between two kinds of knowledge is a bad idea. Its badness stems from the conclusion, which Brandom (takes Hegel to) see(s) as inevitable, that *genuine* propositional knowledge becomes impossible about the domain of representeds. The short reason for this is that, in the two-stage representationalist picture, what is primarily intelligible (i.e. propositionally knowable) are representings: our thoughts, beliefs, and other mental contents. Representeds are intelligible only derivatively as *appearances*, thus never as they are *in-themselves*. There is an unbridgeable epistemic 'gap' between the subject and the representeds when compared to the relation between the subject and her representings, which, almost by definition, downgrades the latter into not-genuine in its status. (Why only *almost* will become apparent below.) What prohibits the knowledge of representeds from being genuine is *not* simply that such knowledge is always *potentially* suspect to error, but because that relation is contrasted to a higher kind of epistemic relation, which, as a necessary condition for the lower kind of knowledge, must always remain 'second-hand', thus susceptible not just to errors but fundamental scepticism as well. The susceptibility to scepticism, in turn, is why Brandom considers the two-stage representational account to be a bad idea.

To remedy the two-stage view, Brandom proposes a 'Genuine Knowledge Condition' (GKC) that should be met by any adequate epistemic and semantic account of the possibility of propositional knowledge about representeds. The term 'genuine' here has a structural meaning: what must be avoided is a setting where there are two classes of propositional knowledge, one primary and incorrigible, the other secondary and corrigible, which automatically makes the secondary knowledge 'ingenuine' and susceptible to scepticism. To specify, it is not merely the bare difference between different kinds of knowing-that, which leads to scepticism, but the fact that one of these is privileged as error-free and explanatorily prior to the second. The first necessary criterion of adequacy, which Brandom thus sets for himself, is that both the knowledge of representings and knowledge of representeds must be semantically equal with each other, at least in the negative sense that the knowledge of representeds is no less genuine than the knowledge of representings.

What makes GKC tricky to fulfil is that Brandom cannot simply reverse the two-stage representational account by taking the knowledge of representeds to be primary and genuine, thus the knowledge of representings to be secondary and ingenuine. Neither does he want to drop representings from the picture altogether. Insofar as we think that 'representation' is in anyway a necessary concept in our endeavour to explain conceptual contentfulness, the fact remains that the knowledge of representings must be in *some* way primary to the knowledge of representeds, simply because the former is how we come to gain the knowledge of the latter. Translated into Fregean terms, semantically, we need the concept of meaning/sense to explain (a) the subject's cognitive 'grasping' of referents and (b) the determination of the referents of expressions.

How, then, is the epistemic equality between the two kinds of propositional knowledge to be attained? Here is where the three other criteria of adequacy come in. The second of these is called 'Intelligibility of Error Condition' (IEC). What this means, briefly, is that semantic error must be made intelligible on *both* sides of the epistemic gap to make it bridgeable; it must be possible to be in error about representings themselves and not just about representeds. The initial difficulty here is to make sense of semantic errors that are not representational in kind since, as we already saw, the knowledge of representings cannot (everywhere) be representational.

The third criterion of adequacy is the 'Mode of Presentation Condition' (MPC). According to MPC, representings 'must be essentially, and not just accidentally, appearances *of* some purported realities. One does not count as properly having grasped an appearance unless one grasps it *as* the appearance *of* something' (ST, 45). The main point here is that senses should not be thought of as entities metaphysically independent of their referents, but rather adverbially as the mode in which the referent appears to the subject. In successfully grasping a determinate referent by an expression, the subject relates not to the appearance as an intermediary entity but to the referent itself, considered in the particular mode.

The fourth, final and arguably the most important condition is the 'Rational Constraint Condition' (RCC). The idea here is that the representational relation must be understood as a normative relation: 'taking or treating something in practice as a

representing is taking or treating it as subject to *normative* assessment as to its *correctness*, in such a way that what thereby counts as represented serves as a *standard* for *assessments* of correctness.' Furthermore, the represented must be such that it provides *reasons* for subjects' use of representings that are about it (ST, 46). Initially, it might seem that RCC is trivial for any semantic theory of representation to fulfil, for surely enough, if my word 'dog' represents dogs, my use of it is to be semantically evaluated in relation to truths about dogs. However, what Brandom has in mind with RCC is anything but a trivial claim:

Reasons are things that can be thought or said: cited *as* reasons, for instance, for an assessment of a representing as correct or incorrect, as amounting to knowledge or error. That is to say that what provides reasons for such assessments must itself, no less than the assessments, be in *conceptual* form. Giving reasons for undertaking a commitment (for instance, to an assessment of correctness or incorrectness) is endorsing a sample piece of reasoning, an inference, in which the premises provide good reasons for the commitment. It is to exhibit premises, the endorsement of which entitles one to the conclusions. So the reasons, no less than what they are reasons for, must be conceptually articulated. (ST, 49)

To summarise, Brandom's thought in the beginning sections of ST goes roughly like this. In order to avoid foundational scepticism about our knowledge of representeds, we must reject the unbridgeable *epistemic* gulf between two kinds of knowledge: of representings and of representeds, respectively. In order to do that, the two kinds of knowledge must be brought to equality by conforming to IEC, in the sense that error in the relevant sense becomes intelligible on both sides of the representational relation, so that neither is to be construed as essentially incorrigible. And in order to do that, we must explain representings and representeds as *semantically* capable of relating to each other per MPC and RCC. Together, what MPC and RCC demand is that if representings must be understood as being (or about) some representeds, and if representeds are to serve as reasons for evaluating the use of representings, then both representings and representeds must be in *conceptual* shape. If only what is conceptual (i.e. assertable) can serve as a reason, and if representeds must serve as reasons for the use of representings to be representeds, they must be in conceptual shape, which, for Brandom, means articulated either in deontic or alethic modal relations (so that there is nothing especially *mental* in conceptual relations).

15.2 Referents as Ideal Senses

The main building blocks of Brandom's semantic approach in ST should be familiar since the core architecture (at least in an outline) is fundamentally continuous with MIE. The starting place is to explain what it is for a subject to treat two commitments as materially incompatible with each other. This is done by explaining what it is for the subject to integrate the incompatibility in her existing constellation of acknowledged commitments by following the three Fundamental Discursive Norms, most importantly, the (intrasubjective) Principle of Rational Rectification. Then, it is argued that implicit in treating two commitments as materially incompatible is a concept of *representational purport*, i.e. taking the two commitments to be *about* the same object. Finally, what it is for representational purport to *succeed* is explained in pragmatist-functionalist terms, as what it is for the subject herself to implicitly distinguish between reality and appearance (or in Hegelian terms, being for-consciousness and being-in-itself). To summarise: from the concept of experience of error, we get the concept of representational purport, from which we get via a pragmatist-functionalist explanation the distinction between appearance and reality.

This trident of commitments is (roughly) as far as Brandom goes already in MIE. The problematic part left somewhat open is how to understand the status of the pragmatist-functionalist distinction between appearance and reality itself. As I argued in Section 10, on the one hand, Brandom's official answer here is that the distinction is *nothing but* structural, that there is no meta-perspective from which to assess the distinction of appearance and reality as such. On the other hand, this formulation appears to contradict the less formal characterisations of semantic objectivity that Brandom offers in MIE, where it is claimed that *objects themselves* can exercise authority over what is claimed about them. But how is that possible, if there is nothing else to semantic objectivity than the structural distinction between what everyone takes to be correct and what is correct, since that formulation omits any mention of objects themselves?

In Section 10, I argued that a part of Brandom's solution for supporting the more robust or substantial notion of representational success, which takes objects themselves as having authority over representings, is conceptual realism. As I see it, one important purpose of ST is to see that idea through by integrating the

pragmatist-functionalist order of explanation with conceptual realism. This is done, in part, by satisfying MPC and RCC. My purpose in this section is to show how that happens, and what implications this has for understanding referents, senses, and their relations.

To begin with, let us see how MPC is to be satisfied. As already mentioned, this part of Brandom's theory in ST preserves a lot of the ideas made explicit already in MIE. Most importantly, what remains the same is the dual pragmatist strategy to (a) explain representational success in terms of representational purport (what it is to be a correct representing in terms of what it is to take or treat something as a correct representing) and (b) explain representational purport in functionalist terms as what it is to acknowledge and rectify incompatible commitments in practice. In the paradigmatic example of ST, the subject first perceives a stick to be straight and then bent when immersed in water. Faced with these subsequent, materially incompatible commitments ('The stick is straight' and 'The stick is bent'), what is it for the subject to treat the commitments as appearances of one and the same object, the stick?

The first part of the solution is to recognise that while the two subsequent observational commitments are said to be something *for* the subject, the very incompatibility of the commitments is said to be something *to* her. In other words, while being-for the subject is an explicit form of presentation (not yet *representation*), being-to the subject denotes an implicit form of presentation. It is the difference in what the subject could express in a propositional judgement (either saying 'The stick is straight' or 'The stick is bent') and what she practically does in rejecting one commitment in favour of the other. The way how these states come to exhibit representational purport is by playing distinct roles in the process of experience (*Erfahrung*), where their normative statuses as correct or incorrect are relative to other commitments acknowledged at each given stage. To follow the simple example, at the first stage, the subject treats the observational commitment 'The stick is bent' as correct, while at the second stage, she treats the subsequent observational commitment 'The stick is straight' as correct. Now she finds herself with two incompatible commitments: they cannot both be correct. At the third stage, the incompatibility is rectified, and a new, unified constellation of commitments is acknowledged:

In the context of collateral beliefs concerning rigidity, what can change the shape of rigid objects, and the relative reliability of visual perception under various conditions, the straight-stick belief is accepted as a *standard* for the assessment of the *correctness* (veridicality) of the bent-stick belief. Because they are incompatible, the latter is rejected as incorrect according to that standard. The bent-stick belief is assessed as responsible to the constellation of commitments that includes the straight-stick belief. All of this is to say that as presented in the straight-stick judgement, the straight stick is performing the *normative* functional office characteristic of the reality represented by some representing: it is practically treated as being, it is *to* consciousness, an authoritative standard for assessments of the correctness of representings that count as about it just in virtue of being responsible to it for such assessments. (ST, 78)

The important point here is that the evaluation between different appearances (or senses) is *internal* to the process of experience considered as a judgement-forming mechanism. The reason why the straight-stick commitment (or judgement) prevails over the bent-stick commitment is because the subject holds certain other background commitments as constant: the standard of correctness here is not the stick itself, but the sense privileged as correct by the background set of commitments (i.e. the sense *qua* which the stick appears straight). The answer to what it is for a commitment to be an *appearance of* something is thus given by its role in the process of experience, where the statuses of (mere) appearance and reality are relative to the background of auxiliary commitments. This answer means that MPC has been, for the essential parts, satisfied, for we understand what it is for a commitment to be an appearance of something, a sense.

It is here precisely where Loeffler's question, encountered first in Section 11.1, returns to relevance. Supposing that we accept the aforementioned account about what it is for a commitment to be an appearance of something and to exert representational purport, why should we think they are appearances of the stick itself and actually succeed in referring to it? Why should we think the commitments and their relative statuses as correct or incorrect reach onto ('transcend') anything beyond the judgement-forming mechanism?

To answer this problem, Brandom proposes that fulfilling RCC suffices. For him, Loeffler's question is naturally interpreted as concerning the *rational status* of the constraints that govern the operations of the judgement-forming mechanism. To be counted as genuinely rational, it is necessary that the mechanism is constrained by how the world really is and not merely, say, by the three internal standards of FDN.

The question then becomes: how can the stick itself provide such a constraint since it is not a part of the judgement-forming mechanism?

Brandom's solution, following Hegel, is to claim that the world itself is conceptually structured, and thus fit for rationally constraining the judgement-forming mechanism. Earlier, I already vaguely gestured towards a way to understand the idea, namely that the two poles of the intentional nexus (on the subjective side, senses, and on the objective side, referents) is given a modal interpretation such that what is subjectively determinate must be articulated in deontic vocabulary, while what is objectively determinate must be articulated in alethic vocabulary.

There is, however, an apparent problem here pointed out by Dean Moyar (2020), which shows that Brandom's problem of semantic objectivity cannot be solved simply by bolting conceptual realism as an independent metaphysical thesis to the up-and-running scorekeeping practice. The problem is with how the various sense- and reference-dependence claims in ST are ultimately supposed to relate to each other. To start off, conceptual realism stands for the claim that alethic modal material incompatibility relations (such as the melting point of copper) are reference-independent from discursive practices. The melting point of copper is what it is independently of discursive practices. On the other hand, objective idealism means 'a *reciprocal* sense-dependence of the concepts articulating the *objective* things and relations and the concepts articulating the *subjective* thoughts and practices of understanding consciousness itself' (ST, 209). These two claims are compatible with each other because they concern different domains: the order of understanding and the order of things, respectively. However, in Chapter 12 of ST, where Brandom expounds on his functionalist account of reference elaborated above, he claims that 'Hegelian referents are expressively ideal senses' (ST, 435). As we just saw above, the straight-stick commitment functions as a standard of assessment for the bent-stick commitment, not due to the former's veridical relation to reality, but due to the subject's auxiliary commitments regarding the behaviour of physical objects, which she holds fixed in the context. But if the status of a commitment as successfully referential is ultimately a functional matter internal to the scorekeeper (and her recognitive community), where does that leave referents *qua* mind-independent alethic relations of material incompatibility? How is the privileged, ideal sense

supposed to relate to the alethic modal fact that is claimed to be reference-independent of discursive practices?

Moyar's proposed solution to this tension, which is fundamentally the same as what MIE already ran into, veers back to interpreting Hegel and thus falls outside my interpretive scope of choice. I think Moyar is right though in that Brandom wants to have both his realism and his idealism, while giving a certain explanatory priority to the latter (2020, 94). In any case, as I have mentioned before, my reading of ST focuses on objectivity on the pragmatic side, where the question is how genuine normative force, as opposed to content, can emerge in the scorekeeping practice; on that account, it is not essential for my purposes to settle how conceptual realism, objective idealism, and conceptual idealism are exactly supposed to fit in the same picture. That being said, since the issue is not completely independent of the problem of reason's sovereignty, it will be useful at this point to give a sketch of an answer to what I believe Brandom's solution looks like.

The key point is the idea that not only our epistemic but also our semantic access to the world is fundamentally perspectival in nature. This is something that the criticised two-step representationalist picture also accounts for, but in a wrong way. Our access is 'perspectival' not in the sense that there is some kind of filter (e.g. Kant's transcendental categories) between the subjects in discursive practices constituted by normative processes and the world of objects as constituted by alethic modal relations. Instead, our practice must literally be able to access the world as it is in-itself, for that is a necessary condition for the practice to be rationally constrained by how the world really is. As Brandom says, Hegelian senses are adverbial *aspects* of their referents. The kicker is that no synchronic time-slice of the practice can access *all* of reality at once, nor exhaust any individual region of it *completely*. Our theoretical commitments can access truths about the world without the Kantian gap, but never the whole truth.

I have now broadly examined two of Brandom's most significant claims in ST, namely conceptual realism and objective idealism. Together, these commitments offer a picture of a world that is conceptually structured, where 'conceptually structured' means being split in two modal classes, deontic and alethic, respectively. The relation between the classes is representational, which Brandom understands

pragmatically as doing something, i.e. engaging in the process of the experience of sorting out incompatible commitments. This is how Brandom makes his metaphysical idealism shake hands with his epistemological rationalism – by the way of inferentialist semantics understood via normative pragmatics.

It is clear that in all the three major domains – metaphysics, epistemology, and semantics – Brandom’s account merits an extensive and focused discussion. However, what I continue to push as the main theme of my criticism homes in on Brandom’s normative pragmatics, which I take to be the key piece of the grand puzzle sketched above. The primary interpretive task remains to explain what exact content Brandom’s core terms of *institution*, *bindingness* and *genuine normative force* have as well as how they all hang together. Already in Section 12, I mentioned that Brandom’s overall response here draws from Hegel’s reciprocal recognition. As we shall later see, it also draws on Hegel’s concept of *Erinnerung*.

The main purpose of Brandom’s response in ST is, I believe, to explain what it is that we do in using the deontic vocabulary of scorekeeping. As I argued in Chapter II in the context of responding to critics of Turner’s anti-normativism, some authors appear to think that it suffices to save genuine normativity simply to give an expressivist spin and to forgo its descriptive ambitions. But this is not where Turner’s counter-transcendentalist argumentation stops: he could (or at least should) agree that the normative vocabulary plays an expressivist role and still hold on to the claim that in describing what it is that we do in using the normative vocabulary, we do not need to appeal to normative vocabulary in order to explain what is going on. From the fact that no descriptive vocabulary can do what normative vocabulary does, it does not follow that there is anything thereby done, which non-normative vocabulary could not adequately describe. To insist otherwise is to commit the transcendentalist fallacy. Something more is required to save genuine normativity (in the form that most normativists understand it) other than metalinguistic expressivism. And that is precisely what Brandom’s interpretation of *Erinnerung* is set to provide.

15.3 Comparing MIE and ST on Acknowledgement and Action

Above, I showed how Brandom's pragmatist-expressivist strategy in ST of explaining key semantic terms, such as representation and propositional content, retain much of MIE's core architecture examined in Chapter II. On the surface, the newest addition is the metaphysical theory of conceptual realism, which aims to complete the picture about how the process of experience may 'incorporate' parts of the external world. However, aside from conceptual realism, there is an important – one could even say radical – break in ST's conception of the process of experience as opposed to Brandom's work more closely tied to MIE's core architecture. The main difference I have in mind concerns the concept of acknowledgement of material incompatibility, which we have seen plays a central role in the process of experience.

As we recall, it is a defining trait of MIE to distinguish between motivational force (which is a psychological and causal phenomenon) and normative force that is irreducible to psychological concepts and causality, although it has its roots in scorekeepers' normative attitudes, which are at least partly in the psychological domain. This shows, first, in that MIE's primitive explanans is the act of attributing a theoretical (i.e. doxastic) commitment, which is completely independent of motivational issues either on the attributer's or the attributee's part. The same goes for the derivative act of acknowledging a commitment, which, although it can be used to express motivational force (MIE, 552), is not primarily a device for expressing motivation but for expressing responsibility over undertaken commitments (practical or theoretical). And to be (practically) committed to do something does not imply being motivated to do it, much less that one will do it. The crucial point is that the normative force, which grips the subject upon the practical acknowledgement of a commitment, is in no way dependent on her motivations for executing the practical commitment: (cf. Brandom 2008, 175). What this means is that there is a kind of logical gap between the subject's acknowledgement (the normative moment) and her actual behaviour keyed to the acknowledgement (the causal moment) by a reliable differential responsive disposition: a gap where Brandom can place the phenomenon of *akrasia* (AR, 95).

Things are very different on the side of ST. Here is a quote to start with:

The cash value of taking it that there is a standard distinguishing cognitive and practical success or correctness from failure or incorrectness is *acting* differently in response to those assessments. *What* one does to distinguish those cases is respond to the error or failure that consists in undertaking incompatible commitments by *revising* one's commitments: withdrawing or adding some, or adjusting the conceptual content one takes some of them to have—which is to say the material incompatibility and consequence relations one takes them to stand in. This is the deep connection between determinate negation as characterizing *relations* of exclusive difference and determinate negation as a principle and motor of *activity*, change, and disruption. The connection between them is normative: incompatibilities make those who acknowledge them responsible for doing something. (ST, 433)

The difference to MIE's conception of acknowledgement is that, in ST, the act of acknowledging two commitments as incompatible is not *only* a passive registration of their status as incompatible but *also* the very act of rectifying the incompatibility. In other words, whereas in MIE, there is logical room between the subject acknowledging an incompatibility and not actually rectifying it; in ST, this gap vanishes, for acknowledging an incompatibility simply means to rectify one's commitments.⁸⁹ The difference might seem superficial, but it has vast ramifications. For one, already in the quote above, Brandom seems to imply that acknowledgement plays a double role: both a normative moment of responsibility and a psychological moment ('a motor of activity') of *causing* the subject to act in certain ways in terms of practical or theoretical reasoning. In other words, the normative and the causal moments that are so carefully distinguished in MIE seem to an extent collapse in ST's understanding of acknowledgement.

Why is the 'collapsed' double role relevant? The reason is the following: MIE's account of normative force, such as responsibility and its acknowledgement, is more or less in line with our ordinary notion of normativity and motivation, i.e. that it is possible to acknowledge one's responsibility without having motivations for acting one way or the other, or even to have motivations opposite to the responsibility, which is just to say that our folk beliefs leave room for akratic behaviour. (Humeans, of course, roughly disagree, but I ignore them here.) However, in ST's conception, this is not possible, at least with acknowledging an incompatibility: the 'cash value' of acknowledging the responsibility to rectify is the very act of rectifying. It is not

⁸⁹ As a side note, perhaps not all cases of rectifying commitments count as registering incompatibilities: sometimes the subject may change her commitment on other grounds.

possible to both acknowledge an incompatibility and not by that token to rectify it. But this immediately raises the question of redundancy: how can acknowledgement *qua* rectifying action not simply collapse to something 'that merely happens', i.e. a psychological phenomenon of motivation not to believe incompatible commitments? If the subject's acknowledgement of an incompatibility is defined by her (attempted) act of rectifying the incompatibility, why would we need to add that this causal moment is, at the same time, a normative one of acknowledging the *responsibility* for rectifying the incompatibility, since it would be a responsibility that the subject could not (intend to) fail to live up to? (Of course, the process of rectification itself may always fail for some reason, which is not to say it was not a genuine attempt to rectify at all.)

The key to the conundrum, we shall see, lies with Brandom's account of recollection, which, in a sense, attempts to reverse MIE's strict differentiation between motivation and normative force to the effect that for rational creatures, motivational force becomes *entwined* (to put up a placeholder term) with normative force. Recollective action can then be understood as individuated by both 'psychological motives' and genuinely normative force.

In contrast to MIE, one radical idea of ST is that the process of recollectively rectifying incompatible commitments is governed by both motivational and normative force. This point notably ties to another major reversal, which ST performs in relation to MIE, namely the explanatory priority of practical and theoretical reasoning. The core strategy of MIE is to understand practical reasoning on the model of theoretical reasoning and perception, so that rational action is understood as a RDRD mechanism responding to the acknowledgement of practical commitments (intentions) by performing the action, at least when all goes well. But in ST:

the model for the retrospective recollective-expressive discernment of the) implicit unity of a course of experience—the development of what things are for consciousness in the direction of what they are in themselves—is to be found on the practical side of intentional action. (ST, 443)

What this means is that, whereas the strategy of MIE is to model intentions (practical commitments) on beliefs (theoretical commitments), the strategy of ST is to model beliefs on intentions. Furthermore, intentions themselves are understood as kinds of

norms, which govern action in the dual sense described above, namely motivationally and normatively:

Normative government [of actions by intentions] has two aspects, one deontic and one alethic. On the deontic side, the intention retrospectively discerned serves as a standard for assessment of the success of each of the subdoings, accordingly as it contributes to the realizing of the intention. On the alethic side, the subdoings rationally reconstructed as a plan structured by the intention are recollected as having been subjunctively sensitive to that intention, in the sense that if the content of that organizing norm *had been* different, the various subdoings *would have been* different. (ST, 444)

If true, this account of the dual nature of intentions as norms entails an explanation of how normative reasons (justifications) can also be explanatory reasons (motivations) by explaining the agent's actions, not simply by what she regarded as correct, but what literally is correct, so that the normative fact of correctness becomes incorporated in the causal explanation of action. That is to say, ultimately in this view, it is not a pseudo-explanation to say (at least in explaining the meaningful use of expressions) that the rational agent acted as she did *because it was the correct thing to do*.

16 ON RECOLLECTION

In Section 14.4, I examined (the outline of) Brandom's formal solution to the problem of reason's sovereignty, the key of which is to combine the attitude of attribution and the status of authority as a privileged combination on which the model of mutual recognition is erected. The substantial issue over the origins of genuine normative force, however, remains the same as with the earlier models: how come that anyone can have the authority to attribute authority to others? If the model of mutual recognition is supposed to solve the problem of reason's sovereignty, it must be self-sustaining, i.e. it must not postpone the question of genuine normativity's origin to some further metaphysical investigation. The way the model can be shown to be self-sustaining is to show how recognition can be *de facto* symmetrical. For in that case, my authority to authorise your attitudes is authorised by your authority to authorise mine. The problem, of course, is to explain how the attributions of authority can 'bootstrap' themselves in this manner, supposing that my attribution cannot be authoritative *before* yours is, nor yours before mine.⁹⁰

The purpose of this first section is to give a rough summary of what I think is the essence of Brandom's solution to the temporal problem, with the later sections delivering the details. An illuminating starting position is Davidson's (1968/2001) claim that the rationalisation of action is a species of causal explanation. What Davidson calls the 'primary reason' for action consists of a pro-attitude, according to which an action of a certain type is somehow valuable and a belief that the action performed is of that type. Rationalisation then explains behaviour if primary reasons can be causes, which Davidson argues they are.

⁹⁰ One might wonder if the simple solution to this 'temporal' problem is that the recognitions must be simultaneous. There are two reasons why simultaneity is not in the cards here, or why Brandom does not even bother mentioning it. First, supposing the truth of the general theory of relativity, there is no such thing as absolute simultaneity, but only simultaneity relative to a frame of reference. Second, even when the relativity of simultaneity is accounted for, simultaneous mutual recognitions are bound to be a relatively rare occurrence, barring vagueness in how precise the simultaneity must be. I am not bringing this discussion into the main text because I think it is clear already on these grounds that the temporal problem is logical or conceptual in nature, not physical or empirical.

Brandom agrees with the idea that rationalisation by primary reasons is a species of causal explanation, and hence that reasons can be causes. However, he differs in ST from Davidson's basically Humean understanding of reasons in opposing the logical difference between motivating and justifying reasons (ST, 554). As we saw in the section above, ST understands the attitude of acknowledging commitments, both as a motivating 'motor of activity' and a normative moment of undertaking responsibility. In other words, whereas Davidson would say that a primary reason means a reason that the agent takes to justify the action in the sense that there was something valuable in it (at least for her), it is a different question (including from the agent's point of view) whether the action was justified in the 'ought to be done, all things being equal' sense. But since for Brandom, the justification for recognitive action, also in the second sense, must originate from the model of mutual recognition itself (i.e. the model must be self-sustaining), he ultimately must collapse the motivating and justifying senses of reason, so that the logical gap between 'ought to be done all things considered' and primary reasons vanishes. Another way to put this would be that for Brandom, at least some primary reasons must be really justified (i.e. justified in the 'ought to be done, all things being equal' sense) in order to be primary reasons.

Collapsing (at least partially) the concepts of motivating and justifying reasons has important consequences for the causal structure of action. Another important Davidsonian (or Hegelian) idea, which Brandom implements, is called the 'essential unity of action', meaning that an action is not a causal chain of three distinct moments (forming the intention, performing the action, the unfolding of the consequences) but rather a single event with no internal causal structure. Of course, virtually any event may be discerned into further constituting events, but Brandom's point that matters here is that for an event that is also an action, the distinction between (i) the intention, (ii) action itself, and (iii) its consequences cannot be drawn in causal terms. While these elements correspond to different descriptions of the same unitary event, what are at stake are not causal descriptions but normative ones (ST, 387-396).

To illustrate, consider an event (chain), such as moving my finger, flipping the light switch, lighting up the room, and alerting the burglar. At what point, according to the 'unitary model', can we start and stop describing the event as an action?

Brandom's answer: at the point where the event can be described as *intentional*. Moreover, if the event is intentional under one description, it counts as an action as a whole. The answer to what makes a description of an event intentional is the availability of the descriptions as roles in premises in practical reasoning.

How is any of this supposed to help solve the temporal problem of how mutual recognition can be *de facto* symmetrical? The key thought is that rationalisation is a kind of recognition. In rationalising B's action, when all goes well, A effectively recognises B's primary reason as explanatory and justificatory of the action. It is explanatory in the subjunctive sense, mentioned in the previous section, that had the primary aim (the thing seen as valuable by B) been different, so would B's action have been. The rationalisation is further justificatory (thus properly cognitive) to the extent that A also accepts B's primary reason as justified, i.e., as a reason which she herself (and anyone else in that position for that matter) ought to endorse.

The final major piece for solving the temporal problem comes with Brandom's claim that by successfully justifying B's action, A also comes to *change* what the action was 'in itself', i.e. the actual content of the action, of what was, in fact, achieved by the action. In other words, the solution comes with roughly the following temporal order:

T₁: B petitions for recognition (justification) by A for action *a*.

T₂: A grants recognition (justification) for *a*.

T₃: A's granting of recognition to *a* becomes justified by the attribution of authority implicit in B's petition for recognition at T₁.

What is going on in this simplified schema is a certain retroactive effect, where A's authority to grant recognition to B's action becomes justified by an *earlier* attribution of authority by B, which comes to have authority itself only *after* A's recognition. It is important to emphasise that at stake is not merely an *appearance* of retroactive effect, but an actual change of the past in view of the future. Equally important is to realise that the change is not causal but normative in nature: it is B's authority to recognise A's authority that changes, or becomes valid, *ex post facto*.

Retroactive justification makes mutual recognition *de facto* symmetrical because the crucial point – who recognises whom *first* – becomes relativised to a perspective, so that in a sense, there is no (absolute) first moment of recognition and thus no need to decide the matter one way or another. From B's perspective, her petition for recognition by A is justified by the authority A has independently of B's attitudes, and vice versa for A's recognition of the petition, which A takes to be justified by B's primary reason itself. So, from the perspective of each, the other has authority first, whereas in reality, the priority of authority *shifts* in the course of the schema.

But now, an obvious question arises: if the priority of authority is relativised to scorekeepers' *perspectives*, how possibly can the resulting retroactive effect not avoid the status of mere appearance? This, we will see, is a vexing point around which Brandom's Always Already argument has to navigate. The basic form in which the argument seeks to solve the temporal problem is by claiming that there is a kind of necessary 'structural' commitment involved in being a discursive being, according to which one is already always committed to recognising the authority of others' petitions for authority. This is shown by arguing that anyone who doubts this thereby commits a pragmatic contradiction where their explicit attitudes are incompatible with their implicit status as a scorekeeper.

In my view, ultimately, there is no avoiding the conclusion that the retroactive effect – and by that token, genuine discursive normativity itself – is at most a mere appearance and not reality. Whether it is, in some sense, a *necessary* appearance is an interesting question that will have to await future work to be fully addressed.

16.1 Introducing *Sittlichkeit* and Modernity

Brandom dedicates practically the whole of Part Three of ST to the temporal problem of *de facto* symmetrical recognition schematised above, though, of course, he has to deal with a lot of other issues as well. Following Hegel, the argument is structured as a sort of an intellectual history between two epochal cultural formations: '*Sittlichkeit*' ('ethical life') and 'Modernity'. In the context of my work, which actively ignores the linkage to Hegel's exegesis, I take it that the most useful way to render these concepts is as metanormative quasi-theories that are somehow

present in the self-reflective attitudes of the discursive community. Consequently, I find it natural to contextualise Brandom's discussion in Part 'Three of ST' as a development on the scorekeeping practice presented in MIE. This first section in particular is meant to examine what changes when the triply perspectival scorekeeping practice becomes 'self-reflective' in the metanormative sense.

Brandom follows Hegel's famous example of Antigone as the paradigm of describing *sittlich* norms in action. The main plot of Sophocles' classic tragedy is about Antigone, daughter of Oedipus, who is torn by two materially incompatible duties. On the one hand, she is tasked to bury her slain brother, Polynices, while on the other, she is forbidden from doing so by Creon, the ruler of Thebes, who considered Polynices his enemy. The conflict counts as a tragedy in the original sense of the word that Antigone is obliged to do the impossible. As a member of the polis, she is obliged to obey Creon by not burying her brother, yet as a member of the family, she is obliged to do just that. In view of *Sittlichkeit*, this is not merely a conflict between personal interests or desires, for the norms that bind Antigone are seen as parts of the natural order, independent from the attitudes of subjects both in content and (alethic and deontic) force.

What is of interest to Brandom in this multi-level conflict is that both Creon and Antigone regard the question of what ought to be done to Polynices' body as settled by *sittlich* norms. This means that both Antigone and Creon implicitly recognise the sensitivity of Polynices' normative status to what is actually done to his body. Should he be buried, that is to make it so that he was a citizen already always. The *deontic necessity* that is involved in the act is *retrospective* in that the actual outcome will *reveal*, in view of the *sittlich* norms, which norm was right all along. For according to *Sittlichkeit*, which takes norms to determine both what is correct and what actually happens (to the attitudes, at least), whatever does happen had to happen necessarily in both the alethic and deontic senses.

Things look very different from the 'Modern' viewpoint. The conflict of Antigone and Creon is not really about norms understood as pieces of natural furniture, but rather about the conflicting attitudes themselves, which merely take each other to answer to some definite, independent norms. The norms 'themselves' do not exercise any kind of force without the active uptake of the attitudes, which are both necessary

and sufficient to determine the contents of the norms. The Modern viewpoint is therefore deeply anti-tragic: the subject is really responsible only to what she takes to be her duty. Modernity in this sense is inaugurated by what Hegel called 'the right of intention', the idea that an action's correctness should primarily be judged by the actor's intentions and not by its actual outcomes (ST, 492).

Let us now compare this analogy to how things stand in the scorekeeping practice. Section 12.7 examined the criticism of the scorekeeping model from the perspective of a Martian observer. The main point that I emphasised, following McDowell, was that if it is not necessary for the Martian observer to describe the practice in genuinely normative terms, then it does not matter for the objectivity of genuine normative force that the scorekeepers themselves have to describe their own practice in these terms from their internal perspective. In order for the problem of reason's sovereignty to be solved, normative force must be graspable by some external standard, which the Martian too would be obliged to recognise. Now, we can initially understand the addition of self-reflexivity to the scorekeeping practice as the question whether it would be possible for the scorekeepers to play the Martian for themselves. That is to say, although they would all agree on the formal validity of incompatibility semantics as a description of how their discursive abilities operated, at least some of the subjects would raise the further question of whether the fundamental discursive norms (FDN) governing their discursive processes were really binding, really in force, or whether they could be understood from outside the practice as an internal effect of the attitudes.

The possibility of 'playing the Martian' can be understood in two senses. In the theoretical sense, some scorekeepers – let us call them 'the Naturalists' – would acknowledge the commitment to some combination of reductionism, eliminativism, or deflationism about the normative vocabulary used to make explicit their discursive commitments. They would thus accept the Martian sentiment that their discursive actions are sufficiently explainable in some causal-dispositional vocabulary on which the normative scorekeeping vocabulary supervenes. In the practical sense, some scorekeepers – let us call them 'the Jesters' – would superficially continue to engage in the scorekeeping practice just as everyone else, but with the provision that they always adopted an *ironic* attitude towards their own scorekeeping. That is to say, every attribution and acknowledgement of a commitment or entitlement would be

accompanied by a special 'as-if' operator that did nothing to affect the contents attributed and acknowledged, but rather neutralised any genuine normative force implied in the commitments. So, when attributing an incompatibility to someone else's commitments, a Jester would treat the subject *as if* they were really obliged to rectify the incompatibility. The Jesters may additionally hold Naturalist theoretical views about their practice, and it would be natural for one belonging to either camp to be a member of the other too, but this is generally not essential.

The counterpart of the Jesters and Naturalists can be fluently called 'the Normativists', which number those scorekeepers who hold opposite views in both the theoretical and practical domains. The crucial thing, however, is not the comparison between these opposing camps within the scorekeeping practice, but rather whether the Martian view is possible for the scorekeepers to adopt on themselves to begin with. This is because Brandom's foundational argument in favour of genuine normativity is predicated on the idea that the Martian perspective is somehow a defective form of semantic self-understanding or self-consciousness. In a word, it is defective because it is alienated. The strength, not to mention validity, of the Always Already argument is thus directly proportional to the nature of 'defectiveness', which will become gradually clearer in the following sections.

We can now tackle Brandom's epoch-defining concepts of *Sittlichkeit* and Modernity, beginning with the former:

Sittlichkeit is then a matter of the *bindingness* ("Gültigkeit") of norms. That is, it concerns the nature of their *force* or practical significance. The Hegelian image is that one is *at home* with *sittlich* norms, one *identifies with* them. They are the medium in which one lives and moves and has one's being. Ultimately, this is a matter of them being a medium of *self-expression*—understood as *constitutive* self-expression. (ST, 473)

Sittlich norms (or rather, *sittlich* meta-attitudes towards norms, but I will often abbreviate this phrasing like Brandom does) are special in that identifying with them means being governed by them in both an alethic and a deontic sense. One is alethically bound by a *sittlich* norm when one's attitudes are subjunctively sensitive to the norm, in that if the norm had been different, so would have the attitude been. One is deontically bound by a *sittlich* norm when one's attitudes are assessable according to their correctness and incorrectness to it. In other words, it is defining of *sittlich* norms that they are (taken to be) both normatively and causally efficacious

in determining the attitudes of the subjects who are bound by and identify with them. To 'identify' with the norms means that one sacrifices (i.e. gives up) one's particular discursive attitudes wherever these are found to be incompatible with the norms. The *sittlich* attitude thus corresponds to the 'Normativist' camp, whereas 'the Moderns' corresponds to the Naturalists and Jesters. As already mentioned, the Naturalists and Jesters do not as such disagree with the bare mechanics of scorekeeping and incompatibility semantics: they only claim that what the subjects sacrifice their attitudes for is not norms as such but only other attitudes (ST, 475), even though it may not appear to be so to the subjects themselves.⁹¹

The key characterising trait of Modernity that sets it apart from *Sittlichkeit* is called 'alienation' (*Entfremdung*). The modern subject is alienated from the norms in that she cannot *identify* with them as the *sittlich* subject could, simply because she does not recognise the norms as having authority in the way *Sittlichkeit* does. However, Modernity does not imply that all norms are given up, but rather that their justification becomes problematised. Accordingly, Brandom understands the modern problem of alienation precisely as a problem of justifying the norms, given that they are not justified by tradition, nature, God, or any other non-attitudinal fact (ST, 501). What makes the issue acute is modernity's implicit commitment to reductive naturalism about norms, i.e. the descriptive thesis, which (roughly) states that all norms are somehow construable as the effects of attitudes. One of the main questions that Brandom pursues in the third part of ST is whether Modernity can successfully combine a normative story about the justification of norms with a descriptive story about how norms are reductively explained as the effects (or products) of attitudes. Doing so would amount to overcoming alienation, i.e. recognising the norms as valid in the sense of identifying with them.

⁹¹ The Naturalist claim that genuine normativity is merely an appearance 'from inside' a discursive practice is, of course, metaphorical and must remain so for the moment. A useful analogy, however, could perhaps be drawn to physics, where it has been known for a while now that (e.g.) the so-called 'Coriolis force' is not a real (i.e. fundamental) force, even though it appears to be so from within a given limited physical system like a carousel or the Earth. Similarly, the Naturalists need not claim that genuine normativity is merely an illusion. The effects may be real enough, but they are not fundamental or *sui generis*, so that Naturalism in this sense can opt either for reductionism or eliminativism.

16.2 The Problem of Alienation: An Initial Look

Brandom states that, for Modernity, 'alienation is at base a pathology of legitimation, undercutting the bindingness of norms' (ST, 501). The purpose of this section is to clarify how and why exactly the pathology of alienation unfolds for Modernity.

To begin with, the problem is initially *descriptive* and metanormative, but it becomes 'pathological' in the *expressive* domain of normativity. (Later on, we shall see that the expressive pathology, the inability of the Jesters to *do* something, turns into an argument as to why the Naturalists should not *believe* what they do.) As we already saw, *Sittlichkeit* stands for a kind of practice where first-order normative claims (i.e. prescriptions about what to wear on Sundays) are justified by the appeal to certain kinds of descriptive claims ('because the Pope said so'). Be it tradition, nature, or religion, whatever the source for *sittlich* norms is, the community accepts certain inferences from descriptive premises to normative conclusions simply because the descriptive premises include a commonly accepted regress-stopper of justification. However, when modernity points out the 'naturalistic fallacy' inherent in the *sittlich* community practises, the justificational link becomes severed to the effect that inferences from the descriptive base will no longer be accepted as legitimate. After the modern turn takes place, potentially all norms (including discursive ones) can be questioned as to their justification and genuineness.

Clearly, a kind of conflict sparks between Modernity and *Sittlichkeit* within the scorekeeping community; the nature of it will become the topic in Section 16.4. For the moment, the focus is on Brandom's claim that once some scorekeepers take the Modern turn and become Naturalists, it becomes impossible for them to *do* something which the normativists can do, namely justify their own normative practices. Note that the Naturalists need not be aware of this incapacity of theirs, for they may continue to *take* certain commitments to be really justified in practice. The point is that even though a Naturalist may retain a 'trustful' (i.e. non-ironic) relation to normative practices, according to Brandom, this *trusting will not succeed in its aim because of the metanormative descriptive theory the Naturalist has endorsed*. To put it in another way, every Naturalist doubles up as a Jester whether they realise this or not, so that the practical and theoretical sides of Modernity are indeed not independent in principle. This observation should also make clear that the 'ironic' as-if attitude of

the Jester does not imply that he never means what he says (his speaker-meaning always contravenes his literal meaning), but rather that what he (or anyone else) says is no different in the normative status from the voicings of a parrot, i.e. it is sufficiently explainable and understandable by causal factors.

It is difficult to understand how exactly a change in the descriptive metanormative theory could precipitate a pathology in the expressive normative theory if one does not keep in mind Brandom's more general views on normativity and expressivism, which he relates to Hegel's concepts of consciousness and self-consciousness in ST. The crucial point is that 'It is an essential, principled part of Hegel's general methodology to understand what is implicit in terms of its explicit expressions—to think of those expressions as essential to the identity of what is implicit' (ST, 501). What consciousness is (implicitly) in-itself depends in part on what it is (explicitly) for-itself. This is, in fact, another radical break from the 'layer-cake' methodology of MIE discussed in Section 9.4. As we recall, the basic hypothesis of the layer-cake picture of rationality was that it is possible to theoretically distinguish the necessary, merely implicit sayings and doings that, together, constitute an autonomous discursive practice from the logically optional, historically late-coming expressive additions of making explicit those core features, such as indexical, deontic, and alethic vocabularies. The layer-cake methodology thus presupposes that what is implicit is intelligible, at least from the theoretician's perspective, apart from the possibility of its explicit expression by the scorekeepers themselves.

The dual nature of alienation becomes most clear In Chapter 14 of ST, where Brandom discusses both its practical and theoretical dimensions (Hegel's 'actual' and 'pure' consciousnesses). Alienation appears in both dimensions by repeating a certain substructure, an incompatibility between two structural elements. On the practical side, these are called (in Hegel's allegorical terms) 'Wealth' (*Reichtum*) and 'State Power' (*Staatsmacht*); together, they concern the individuation of actions. Wealth represents the subjective side, which individuates the essential content of actions by the motives and intentions of the agent, in two senses. In the normative sense, the subject is held responsible only for what they intended to do, not the actual consequences of their action, while in the explanatory sense, the action is always explained by reference to idiosyncratic motives, i.e. as advancing the agent's acknowledged preferences somehow. In contrast, State Power represents a

consequentialist normative stance on actions (always holding the agent responsible for what actually happened regardless of intentions) and a 'universalist' stance on their explanation, which means that the action is always explained by reference to some norm, which the agent (presumably) intended to follow.

The relations of Wealth and State Power are alienated to the extent that their normative and explanatory stances are seen as materially incompatible with each other. That is to say, both Wealth and State Power accept that either the agent is to be held responsible for what they intended or for what they actually achieved, but never both; and likewise, in terms of explaining their action, which can only have one cause properly speaking. (To clarify, in the explanatory sense, both sides agree that it is the agent's will that caused the action, yet they differ in whether this will was motivated by particular and idiosyncratic or universal reasons.) In other words, neither side recognises the other; hence, the individual and the community will be locked in perpetual conflict.

I shall postpone the more detailed discussion of the practical side of alienation to Section 16.3. To finish the initial look at alienation, I shall explain its theoretical side. Here, the incompatibility holds between 'Faith' and 'Enlightenment', which can be considered as (quasi-)theories of (meta)normativity (ST, 523). I already outlined the controversy above as a certain change in the descriptive metatheory of a community, where a given premise base of regress-stoppers (tradition, etc.) comes to lose their collectively accepted status as grounds for normative claims. At the same time, the new purported base that consists only of attitudes cannot serve the same role as a justificational regress-stopper, for no one in the community is no longer able to determine or infallibly discern the correct normative claims by their attitudes. The historical community to which this story is (roughly) supposed to fit is, of course, the society of Estates as reformed by Enlightenment republicanism. As such, the Enlightenment criticism of Faith consists of three major points:

Ontological: Belief in supernatural sources of normativity is categorically false because there is nothing aside from attitudes that could ground normativity.

Epistemic: Even if there was a supernatural source of normativity, the actual epistemic justifications for deciphering it are epistemically insufficient.

Moral: Together the ontological and epistemic shortcomings are used for immoral practices, such as the divine right of kings and ecumenical privileges, which really are just abuses of political power and legitimation. (ST, 526–527)

The first two points amount to attributing to Faith a cognitive error, while the third stands for a recognitive one. In a nutshell, from Enlightenment's point of view, Faith is merely an erroneous theory that essentially contains propositions about entities that do not really exist. However, as already noted, while Enlightenment is arguably correct in its cognitive criticism of Faith, the price of its criticism is that it loses what Faith had achieved, namely a community of trust:

The second part of this passage puts three requirements for an attitude to count as trust. The trusting one must recognize her own being-for-self, her own self-conception, in the trusted one; the trusting one must correctly take it that self-conception is acknowledged by the trusted one; and the trusted one must correctly take it that that self-conception is acknowledged by the trusting one also as her own. The first part of the passage says that when those conditions are met, the trusting individual counts as identifying with the trusted individual. (ST, 529, reference to *Phenomenology of Spirit* omitted)

Faith is wrong in its cognitive attitudes, misunderstanding its object and its relation to that object. But it succeeds with its recognitive practices, creating a community of trust. Enlightenment is right in its cognitive attitudes, correctly seeing that the normativity both are concerned with is not something independent of our attitudes and activities. But it fails on the recognitive, practical side. (ST, 534)

The cognitive error of Faith is more or less clear, but what exactly does it mean to say that Enlightenment fails in practice to create a community with determinate norms? Brandom gestures towards the historical event of *la Terreur* as an example, but it is not clear in what sense the example is to be taken. To simplify, we could say that in the 'sociological' sense, a community fails if there is no sufficient level of trust among its members, where 'trust' is understood according to the three conditions Brandom names above. So, in the ideal situation, every subject correctly (i.e. successfully) recognises their self-conception (being-for-self) in every other member, and every member correctly recognises the identification. It is important to note that what the subjects identify with (their 'self-conception') in the others is their sacrifice for the common norm, but not this norm or the other as such. Each identifies with the others' 'universal' half, for which the sacrifice stands.

What the dyadic trusting relation essentially demands is mediation by a norm with which both subjects can identify in the sense of being willing to sacrifice their particular attitudes for it in practice. The practical problem of Enlightenment is that if there are no norms but only attitudes, then trust in this scheme becomes impossible. True, the subjects can still identify with each other's attitudes directly in the sense of sacrificing their own for them in practice, but since every attitude is *prima facie* an equally good guide to what ought to be done, practically the resolution of conflicts becomes difficult and theoretically impossible.⁹² Furthermore, the trusting community in the proper sense demands not just an agreement on normative ends, but also an agreement on the reasons for those ends. Two subjects may agree on what ought to be done purely for instrumental reasons that are supported by two very different preference sets, and in this sense, 'trust' each other, even when they may recognise that the other's reasons for agreeing are not their own. Trust, in the proper sense, requires that the subjects recognise the *same* norm as binding, applicable, and justified in the same cases.

However, even when the aforementioned conditions are met, the result might still fall short of the practical agreement because the subjects may have different interpretations on what the universal norm dictates in a particular case. This is effectively why Brandom thinks Kant's solution to the pathology of alienation is insufficient. What Hegel calls 'Morality' (*Morälität*)

seeks to combine the universal applicability of moral principles (the consequences of the applicability of a rule) with their origin and validation in the free commitment of an independent individual agent to the principles as universally binding (the grounds of the applicability of a rule). Treating a principle as universal in this sense is committing oneself to accept appropriateness of appeal to that principle by *anyone* in justification, challenge, and appraisal of justifications of performances generally. (ST, 539)

According to his scheme (1) principles genuinely constrain individual actions, which are what they are appraised as according to such principles; (2) performances are actions only as so constrained; and (3) there are no (non-formal) facts about what principles are valid apart from the facts about what principles people take to be valid by endorsing or appropriating them—that is, by committing themselves to their

⁹² I cannot here discuss trust from a broader comparative perspective, but it will help to mention two delineations. First, it is clear that Brandom is not talking about trust as a private relation between individuals; for him, at stake is primarily the public relation. Second, it seems to be implicit in his discussion that a purely psychological treatment of trust would be inadequate at least in the public sphere.

validity. These are precisely the elements required for alienation to be overcome. (ST, 541)

In short, Kant's solution that appeals to the categorical imperative as a necessary and a sufficient constraint on rational action only offers a formal solution to the problem of trust and hence the pathology of alienation, the reason being that the imperative cannot guarantee uniform interpretations. Another failed solution which Brandom considers is called by Hegel 'conscientious self-consciousness', which understands the appropriateness of actions to bottom out in the subject's sincere acknowledgements of what they thought was their duty in the particular case. The essential problem here is that if both the bindingness and the content of norms is determined by one's sincere avowal of conscience, both become circular by failing to meet the is-seems distinction. If the content of the norm by which the action is to be judged can be determined solely by the sincere avowal of the agent, then it will become impossible for the agent herself to be wrong about this content itself (ST, 545).

To summarise this section, Brandom's argument for how the Naturalist descriptive metasemantic theory results in alienation as an expressive pathology traverses a certain intellectual history. We start with a *sittlich* community, where the norms are seen as objective features of the world, the task of attitudes being to conform to them in alethic and deontic senses. What follows is a theoretical revolution resulting in Modernity, where at least some scorekeepers relinquish *Sittlichkeit* in favour of a Naturalist view about normativity, where the force and content of norms are fully determined only by attitudes, and in where *sittlich* attitudes are interpreted as an erroneous theory with morally bankrupt undertones. Consequently, although the Naturalist may continue to hold self-consciously faithful normative attitudes in practice, i.e. they refrain from identifying as Jesters, as a point of fact, they all become Jesters. What the pathology of alienation means is that although the Naturalist may invent other forms of self-consciousness to justify their normative attitudes (for example, by the way of Morality or conscientiousness), none of these succeed in conferring determinate contents on norms and thus fail to meet the is-seems distinction. As a result, no Naturalist normative attitude is really justified in view of the metanormative descriptive theory they endorse; and hence, they fail to institute a community of trust.

16.3 Overcoming Alienation I: Introducing the *Kammerdiener*

The above discussion offered a rough juxtaposition between two metanormative theories that are the basis of two contrasting camps within the scorekeeping community. Alienation is initially understood as a pathology of only one of these, namely Modernity's, which fails in the practical dimension of instituting a community of trust with determinate, attitude-transcendent contents. Yet, this is only one side of alienation: in fact, *Sittlichkeit* too is alienated, but in a different sense than Modernity. The alienation of *Sittlichkeit* is theoretical in nature, which is to say it has a false theory about the norms, seeing them as 'part of the natural or supernatural furniture' of the world and independent of the attitudes (ST, 533). But unlike Modernity (i.e. Enlightenment) claims, Faith is not simply an erroneous theory; its cognitive error is not so much about the object of knowledge as the relation between the subjects and the objects. Essentially, what Faith misunderstands is its own contribution to creating the norms (their content and force) with which the community members identify; its alienation is fetishist in nature.

The dual nature of alienation (i.e. its practical/theoretical and recognitive/cognitive sides) means that in order to 'overcome' alienation, it is not sufficient to merely settle the debate either in favour of Faith or Enlightenment. Rather, a new, third synthesising position is needed, fusing the correct parts from both metanormative theories. Doing so would amount to explaining how the community can institute a normative standard that solves the problem of reason's sovereignty.

It is somewhat telling that Brandom's main strategy for forging the synthesis proceeds in the first instance by seeking to refute the Modern understanding of normativity by arguing that it is, in a way, incoherent. His main objection, which will be the focus of this section, is that Modernity necessarily involves a kind of practical failure or pragmatic contradiction; and hence, its word on norms cannot be final. The failure, as already sketched, is the inability to create a community with determinate norms (both content and force). This is a problem because, in Brandom's view, the very possibility of propositionally contentful thought and talk presumes that the subject takes some commitments to be in force and have content independent of her own attitudes. This counts as a 'pragmatic' contradiction because it is not a contradiction in what is said (between two propositions), but between what

is said and the saying of it. Somewhat like the utterance 'I'm a humble person' is not something a truly humble person should (or in a strict sense, could) say, Brandom's idea is that Modernity cannot claim that all norms are attitude-immanent without presuming that some norms are in force independently of the speaker's (or anyone else's) attitudes.

In a rough outline, then, Brandom's strategy is transcendentalist in the sense that it attempts to refute the sceptic by showing that her position is pragmatically self-contradictory. As we shall see, a lot will turn on the point of what exactly the 'pragmatic' self-contradiction amounts to. But to begin with, it will be helpful to get a better look at the sceptic's position as Brandom construes it under the name '*Kammerdiener*'.

The *Kammerdiener* (which is German for a kind of valet) is one of Hegel's allegorical figures, whose contrast is 'the hero':

no man is a hero to his valet, not because the man is not a hero, but because the valet is a valet. (ST, 551)

In allegorical terms, the hero means the subject who performs her duty because and only because it is her duty (i.e. what the norms say ought to be done or what is correct), whereas the valet means the subject who interprets every action as ultimately motivated by 'base' desires, idiosyncratic ends, and personal preferences. In theoretical terms, Brandom characterises the *Kammerdiener* as subsuming three related claims about norms, attitudes, and actions:

Causal idleness. If we understand norms as principles stating what is correct, true prescriptions about what ought to/may (not) be done, or values as to what is good, they cannot causally explain why the subject acted according to what is correct, ought to be done, or what is good. A sufficient explanation of action is what the subjects believed to be (or otherwise practically treated as) correct, prescribed, or good. Only normative attitudes, not what the attitudes are (allegedly) about, (partially) explains why the subject acted as they did.

Error theory. Insofar as normative attitudes have cognitive, propositional content, the contents fail to refer or represent anything in the world: the attitudes are necessarily false. Thus, the truth of the attitudes is neither necessary nor sufficient to explain why subjects have normative attitudes or why their cognitive contents are what they are.

Reductive naturalism. All talk of norms and normativity as well as the actual contents of normative attitudes can be explained in terms of some causal, non-normative

vocabulary. In particular, the contents can be explained in terms of naturalised genealogy, where the attitudes are seen as causal adaptations to the environment. (ST, 561)

As already seen, Brandom's strategy is not simply to refute the *Kammerdiener's* position, but to integrate it into a synthesised understanding of norms and attitudes. The crux of the integration is the third element of the *Kammerdiener's* sceptical position, the genealogical naturalist reductive view of normative attitudes and hence, the norms as well. The main sceptical claim here is that no attitude is really subjunctively sensitive to the norms themselves, but only to some contingent causal factors in the subject's past, most importantly, the attitudes of other subjects. (Compare this to Davidson's primary reasons: it is not the valuable feature itself that explains action, but the belief that the action performed had this feature.) The idea is that if it can be shown, for any normative attitude, that it has the content that it does only because of the contingent causal factors in the subject's history, the attitude cannot be *rational*; that is, it cannot be justified by norms. Since the idea that an attitude could be rational simply by according with (or conforming to) a norm regardless of how (or whether) the norm is a cause of the attitude is refuted by Modernity's contention that there is no normative force independent of attitudes, the only way in which norms may justify attitudes is if they are somehow causally responsible for their content. Moreover, if the attitudes' contents are not constrained by norms, the determinate content of the norms themselves becomes a question, supposing that the content must come from attitudes. 'This project,' Brandom claims, 'is to provide a response to Kant's Third Antinomy—the challenge to integrate reasons and causes' (ST, 558).⁹³

The reductive naturalist, causal-genealogical argument against normative force can be best understood in contraposition to the historical model of content-determination discussed in Section 9.3. The three judges' model posits that the application of norms in judgements is, at the same time, the process of their determination, such that the present judge must appeal to the previous applications in order to petition recognition from the future for her own present application. What the *Kammerdiener* essentially claims is that in order to understand this process,

⁹³ Note that what Brandom has in mind is integration in a stronger sense than what Davidson argued for: the primary reasons must also be really justified in order to be causes for Brandom.

it is not necessary to postulate that all judges are, in fact, subjunctively sensitive in their applications to a genuine norm that runs through their shared history. Rather, each decision of the application can be sufficiently explained by the appeal to contingent local causes that surround it, or in Davidsonian terms, to the belief component of primary reasons.

There are important divisions within the class of contingent causes. For example, since the content of many concepts depends on what language-entry and exit moves are available to the judges, it is in a sense contingent that they can apply, e.g. colour words, as they do. For example, the content of 'ultraviolet' would be different if humans could not naturally perceive it. Moreover, the applications of a word like 'dog' are contingent on the fact that there are dogs in the world. Aside from this sense of contingency, which the Normativist also accepts as effective in determining what concepts are available and what their contents are, the focus class of contingency covers causes, such as what the judge had for breakfast, the kind of legal literature they have been reading, their political views, etc. If such parochial causes are sufficient to determine every individual application, the *Kammerdiener* claims, we need not postulate an extra-causal item (the norm) to which the judge's application would be subjunctively sensitive. And if no such norm is required for explaining the applications, it is neither needed to explain the content of the concept being applied. Rather, the content just consists of (i) the causal history of the concept, (ii) the local circumstances that led the concept to being applied as it was, and (iii) the actual consequences of those applications. (This is, in effect, an extreme version of the historical, finitist account of conceptual contents, which I argued is the result of Kripkenstein's sceptical challenge; I will return to this connection in the final section of this chapter.)

16.4 Overcoming Alienation II: Four Meta-Meta-Attitudes

We have now seen how Part 3 of ST proceeds as a dialogue aiming to synthesise two normative meta-attitudes amounting to proto-theories within the scorekeeping community. In the last two chapters (XV and XVI), Brandom attempts to forge the synthesis by introducing four different ways to understand the conflict between Normativist (or *Edelmütig*/magnanimous) and Naturalist (or *Niederträchtig*) meta-

attitudes; hence the 'meta-meta-attitudes'. The meta-meta-attitudes are ordered along two dimensions and can be displayed in a two-by-two table (of my design):

Table 2. Four meta-meta-attitudes

	Objective	Subjective
Cognitive	1	2
Recognitive	4	3

In the first instance, the nature of the conflict is understood as cognitive and objective, which means that 'the issue is put in a box with the question of whether there are leprechauns, and whether there is a bird in the bush' (ST, 570). In the second instance, the conflict is still deemed cognitive in nature, yet subjective in the sense that while it is incoherent for one subject to adopt both *Edelmütig* and *Niederträchtig* attitudes at the same time, two different subjects doing so will not be in a fundamental cognitive disagreement. In the subjective-cognitive (meta-meta-) view, no single subject can be entitled to both meta-attitudes at the same time, but two different subjects can be entitled to each meta-attitude separately without one necessarily being wrong. This is possible because the cognitive-subjective meta-meta-attitudes amount to adopting a 'stance' on the conflict. Brandom does not name Dennett in this context, but I think it is clear that this option is close to his famously proposed stance stance in understanding intentionality. One can adopt three kinds of stances towards the computer to explain its behaviour: the physical stance, the design stance, and the intentional stance. Different stances come with different pragmatic interests and appropriateness, which still counts as a broadly cognitive, descriptive orientation to the conflict, though one that need not result in the exclusion of the other side.

The conflict in the first, cognitive-objective instance pivots on the idea that there is a single vocabulary in which one can descriptively exhaust the complete 'catalogue of the furniture of the universe' (ST, 570). If there is such a vocabulary, then either there are norms on the list or not (ontological sense) and one can either possess

knowledge of them or not (epistemic sense). The second, cognitive-subjective construal of the conflict introduces a finer distinction in the meaning of a 'complete description of the universe'. Instead of one all-encompassing vocabulary, there can be several different vocabularies, each of which is 'explanatorily complete' in the sense that its domain of phenomena can be sufficiently described using only that vocabulary. For example, a naturalist-physicalist vocabulary is explanatorily complete if all natural-physical phenomena can be sufficiently described in naturalist-physicalist terms. (To avoid a tautology, I assume there is some way to individuate the domains in some other way than using the vocabulary for it.) Insofar as one does not insist that one vocabulary must be explanatorily more basic than all others, in the sense that what can be said in any other vocabulary can be somehow translated into the privileged vocabulary but not vice versa, the meta-meta-conflict vanishes.

The two cognitive combinations both imply a certain 'rejection of a community', according to Brandom. In the objective sense, the meta-attitudes are considered contradictory, while in the subjective sense, they are merely incoherent if adopted by one subject; either way, the subjects are tasked to choose between them. Action can be described either in the normative *Edelmütig* vocabulary or in the naturalist *Niederträchtig* vocabulary, but not in both at the same time or from the same cognitive perspective. The dramatic sounding 'rejection of community' means that there can be no pragmatic agreement between the two sides, which leads Brandom to argue that in a more profound sense, the conflict is not, after all, cognitive but recognitive in nature (ST, 574).

In the third instance, the recognitive dimension is given a subjective interpretation that corresponds to the stance position. What the Enlightened *Kammerdiener* does in rejecting a community with Faith is to identify with the Naturalist side in the cognitive position. The identification happens by sacrificing a certain kind of self-consciousness of oneself, namely as bound by genuine norms (ST, 575). In order to be formally coherent, the *Kammerdiener* must also understand his own actions as something motivated only by idiosyncratic, one could say unfaithful, motives that do not really aim to satisfy any norms, including linguistic norms. 'What the *Kammerdiener* is doing by adopting the *niederträchtig* recognitive stance is making his own and others' performances and practices into something that is unintelligible as *discursive*' (ST, 576). In a word, the recognitive-subjective meta-meta-view amounts to claiming that

it is up to the subject whether he by his self-understanding makes herself a normative or merely a natural being.

The difference between the objective and subjective cognitive positions mirrors the one between their cognitive counterparts in that the choice is construed as exclusive or inclusive. In the objective sense, the choice is exclusive, understood as a point of truth about what either is or is not the case. In the subjective sense, the conflict is not about truth but about two different, *prima facie* equally good pragmatic options. However, the eventual culmination of the four meta-meta-attitudes is the claim that, in fact, the *Kammerdiener's* cognitive position is pragmatically incoherent. Once we understand the metanormative conflict in its more fundamental cognitive dimension, we are also supposed to realise that the objective construal is the correct one, and that under this construal, the *niederträchtig* meta-attitude is in a way self-defeating:

Even the *Kammerdiener* and his resolutely reductive naturalist generalization offer contentful *accounts* of our doings (performances and attitudes), accounts that aim to satisfy the distinctive standards of intelligibility, adequacy, and correctness to which they hold themselves. *If* the determinate contentfulness of the thought and intentions even of the *niederträchtig* is in fact intelligible *only* from an *edelmütig* perspective, *then* anyone who in practice treats what he is doing as judging and acting is implicitly committed thereby to *Edelmütigkeit*. (ST, 577)

If you are implicitly committed to something which you explicitly deny, you contradict yourself in the pragmatic sense. In the case where the implicit commitment is, in a sense, necessary for being a discursive being, what you say contradicts the very possibility of your saying it.

Before proceeding, recall the discussion in Section 12.7, where I claimed that insofar as Brandom wishes to hold on to the idea of genuine discursive normativity, it is not enough to argue that the scorekeeping practice is merely sufficient (and not necessary) to confer propositional contents on expressions. The reason for this claim resurfaces here, for the strategy of the Always Already argument against the *Kammerdiener* precisely is to argue that undertaking genuine discursive norms is a necessary condition for expressing propositionally contentful utterances (and hence that denying that anyone ever undertakes genuine commitments leads to the pragmatic contradiction). Since the claim that the scorekeeping practice is merely

sufficient to confer propositional contents would allow the possibility that our actual linguistic practices, no lesser than the behaviour of the dog with the bone, could be sufficiently explained by non-normative, causal, and dispositional means, it follows that the sufficiency claim is not enough to support the strategy Brandom opts for in defence of genuine discursive normativity in ST. I claim that the situation is largely the same already in MIE, which is thus committed to the necessity of the scorekeeping practice in conferring propositional contents on expressions.

16.5 An Interlude Summary: Alienation and the Problem of Finitude

We are now in a position to summarise the main points of Brandom's treatment of alienation in Chapters 13, 14, and 15 of ST. The account begins with the *attitude* of *Niederträchtigkeit*, which stands for an identification with the 'disparity that action and consciousness involve'. The *Kammerdiener*, who adopts the *niederträchtig* attitudes towards the hero, thereby makes the judgement that the hero is not really 'a hero', i.e. her action (paradigmatically, an application of a concept) was not motivated by (subjunctively sensitive to) *only* the norm, which she acknowledged as the reason for the action/application. Instead, the *Kammerdiener* explains the hero's action as being motivated by various idiosyncratic motives and contingent circumstances in the subjunctive sense that, had enough of these been different, so would the action have been. Transposed into more focused terms, the *Kammerdiener*, as a reductive naturalist, claims that in applying concepts, one's discursive attitudes about what is correct are never *solely* sensitive to the relevant semantic norms but are also affected by various contingent, causal circumstances – most importantly, her attitudes, to the extent that it is the attitudes alone that are sufficient for explaining the applications.

Being widespread enough, *Niederträchtigkeit* leads to alienation as a *recognitive structure* characteristic of modernity, which, in my simplified model, means that the Jesters and Naturalists have decisively outnumbered the Normativists in the discursive community. The alienated structure is defective because it is *asymmetrical*, which means that 'the norms that are applied by the people who are deliberating what to do and justifying what they are doing are not the same norms that are applied by the people attributing those doings and assessing those justifications' (ST, 585–586).

This is the case with 'Moralität' and 'Conscientious consciousness', as explained above.

Brandom observes that the various ways in which modernity represents a defective recognitive structure mutually share a form, which he names 'Mastery'. The core flaw of Mastery is to identify normative statuses with some normative attitudes, or in other words, to violate the distinction between what is right or correct and what is taken to be right or correct: 'If norms are simply identified with normative attitudes (what is correct with what is taken to be correct), the latter become unintelligible too (cf. the 'fallacy of lost contrast')' (ST, 586). The *Kammerdiener*, or what I have called the Jester, judges that the hero's action is not really subjunctively sensitive only to the norm, which she acknowledges as binding. But in doing this, the *Kammerdiener* must himself invest his authority into the truth of this evaluation, and by that token, recognise some standard other than his own attitudes that settles the truth of the matter, or else lose the very idea of determinate conceptual content that presupposes the 'is—seems' distinction. This forces the *Kammerdiener* into a defective form of self-understanding or self-consciousness: the standard that he ascribes to others cannot be the same which he simultaneously, as a judger, acknowledges himself.

Given these ingredients, Brandom wants to argue for (at least) three major claims. The first is the rejection (or rather, an adjustment) of absolute, or Fregean, determinacy: the shift from *Verstand* to *Vernunft*. The second is the idea of 'semantic justification', which means that the contents of our concepts are not completely determined by causal contingency, which he takes to be incompatible with the idea that they could be *rationally* determined. The third is the Always Already argument, which states that, as discursive subjects, we are indeed 'already always' committed to adopting *magnanimous* and trustful normative attitudes, which is supposed to vindicate semantic justification and normativity.

In what follows, my main purpose is to criticise the second and third claims, but before that, it will be useful to see that I am, in a way, in agreement with the major claim number one, i.e. the rejection of absolute determinacy. The purpose of this section is to show that my advocacy of meaning finitism in Chapter I is broadly in line with Brandom's criticism of Mastery and the *Verstand* meta-meta-view of conceptual determination. Doing this will, hopefully, bring to the reader's mind the

adage that one philosopher's *modus ponens* is another's *modus tollens*. For while I agree with the conclusion that absolute determinacy is not possible for finite speakers, I think this also gives reason to think that the juxtaposition between 'causal' and 'rational' determination of concepts is in fact wrong – at least in the way Brandom thinks about it, and for that reason, the notion of 'semantic justification' becomes collateral damage of absolute determinacy's rejection. From this, it is then a short step to refuting the Always Already argument as well.

Before that, let us focus on the first major claim. As with most issues, the rejection of absolute determinacy has, for Brandom, a dual nature: 'pragmatic' and 'semantic'. The picture is more familiar on the semantic side, which concerns the question of how determinate the semantic values (e.g. extensions or inferential roles) need to be for the associated expressions to be properly called meaningful at all. The clearest presentation of absolute ('Fregean') determinacy goes as follows⁹⁴:

Fregean senses are required to determine classes of referents whose boundaries are *sharp*, *fixed*, and *complete*. To say that they are *sharp* is to say that it is impossible for any possible object to fall *partially* in the class determined by the sense (excluded middle), or both to fall in it and to fall outside it (non-contradiction). To say that the referents are *fixed* is to say that the boundaries of the class of referents determined by the sense do not change. (Which sense a given sign expresses may change, if the use of the sign changes, but the senses themselves do not change.) To say that the boundaries of the class of referents is *complete* is to say that the sense determines a *partition* of the possible candidates: *every* particular is classified by the sense either as falling under the concept it determines, or as not falling under it (excluded middle). This is Fregean determinateness, or determinateness in the Fregean sense. (ST, 429)

To be clear, Brandom's rejection of absolute determinacy is only partial in that he thinks (that Hegel thought) that 'an intelligible story can be told according to which what we do, paradigmatically our use of linguistic expressions, gives us access to conceptual contents that are determinate in the Fregean sense.' (ST, 430). The shift from *Verstand* to *Vernunft* is supposed to change how we think about *grasping* and acquiring our concepts, not their determinate nature as such. However, Brandom's thought here is nonetheless broadly in line with Chapter I of my work, where we both reject the idea that, at any given *actual moment*, it would be determinate whether

⁹⁴ Fregean determinateness is not, of course, strictly identical with absolute determinacy, as I discussed the concept in Chapter I. But I think that for all relevant purposes, it is close enough. Accordingly, I will talk about them interchangeably throughout the rest of the work.

Jones is governed by the addition function or the quaddition function. As I understand him, Brandom would instead say that the concept that Jones is in fact governed by becomes more determined over time (if all goes well), though there need not be any time where the concept would be 'absolutely finished'. The important adjustment, which Brandom proposes to Fregean determinateness, is that it is not an either-or question whether Jones is an adder or a quadder in a given context. Rather, we should think of conceptual determination 'stereoscopically', as something which can be looked at both retrospectively and prospectively.

My focus now, however, is on disagreement as opposed to agreement. Turning to the pragmatic half of rejecting absolute determinacy, which, we recall, has to do with the normative force governing concept use, Brandom's starting point is Mastery's insistence that any action can be genuinely governed by norms only if the subject's attitudes are subjunctively sensitive *only* to the norm she acknowledges as binding. What is the *Kammerdiener's* master argument for this claim?

The reasoning goes like this. For any actual act of applying concepts, there are, in principle, two ways to describe what happens. According to what we might call the 'normative description', the subject's attitude and application are subjunctively sensitive to the semantic norm which the subject acknowledges. Simply put, the subject applies the concept as she does because she judges that it is the 'semantically correct' thing to do in this instance. According to the alternative 'causal and contingent' description, the subject's attitude is not subjunctively sensitive to the norm she acknowledges as binding, but rather, it is subjunctively sensitive to some contingent causal features of the context, which explain why she applied the concept as she did, together with her attitudes. Both kinds of descriptions, which are mutually incompatible, are always in principle available for any actual act of applying concepts. Since 'norms are actually efficacious only via attitudes, it is always possible to see the agents as sensitive only to their own attitudes', which, together with certain contingent contextual features of the case, suffices to explain the application as something which merely happens, not what ought to happen (ST, 589).

At this point, the *Kammerdiener's* reasoning may appear as only posing an epistemic problem for the hero. Perhaps it is true that it is impossible for the hero to prove or to know, either in some particular case, or in general, that her attitudes were

subjunctively sensitive only to the norm. But that by itself does not mean there are no heroes. However, the *Kammerdiener's* complaint goes deeper than that, as we can see by comparing it to the problem of finitude posed by Kripkenstein's sceptic. These cases are analogous in the sense that in both, we are tasked to determine why one of two mutually incompatible descriptions of an act of applying a concept is true while the other is false. The important point is this: both the *Kammerdiener* and the sceptic claim that there is no fact about the subject's (normative or dispositional) attitudes, which could determine which description is true. The reason why the subject's attitude cannot determine this is found in the is–seems distinction: if the subject's own attitude suffices to determine whether she really was subjunctively sensitive to the norm, we lose the contrast between what is taken as correct and what is correct. Since it is a part of Brandom's background methodology that determinate conceptual norms exist only in virtue of normative attitudes (i.e. phenomenalism about norms), the result is the same as with the extended argument of Kripkenstein's sceptic, which claims that nothing beyond subjective attitudes can determine which description is true; the result being, there are no determinate conceptual norms governing applications – only contingent causes and subjective attitudes.

And here, we come to the point where Brandom sticks to a *modus ponens* reading of the situation, whereas I would prefer switching over to *modus tollens*. We have rejected absolute determinacy (at least in the form represented by the *Verstand* conception, which I think is analogous to Meaning Determinism), and by that token, cease to seek some fact about the subject that could settle, in any actual context, whether she is governed by the norm of addition or the norm of quaddition. But Brandom thinks that endorsing meaning finitism would eventually amount to a kind of 'semantic nihilism', and worse, the loss of human freedom in the Kantian sense of the ability to bind ourselves with norms. If the application of concepts is ever *rational*, it must be *justified*, and merely causal-contingent facts are insufficient to yield such justifications because they are unfit to be *cited as reasons* for the application, even if they provide true reductive explanations of actions. Moreover, since Brandom ultimately also wants to hold on to something like the Fregean standard of semantic determination, albeit given the *Vernunft* spin, he cannot be satisfied with finitism. Even if we cannot *find* facts to determine the correctness of applications, at least in some cases, such facts can be *made*.

So, how is the trick to be done? As I have already stated, the key piece is to refute the *Kammerdiener* (and by extension, Kripkenstein's sceptic) with the Always Already argument, according to which the *Kammerdiener* (and the Jesters and the Naturalists) commit a kind of pragmatic contradiction, which then leads to a faulty self-consciousness. It is the task of the next subsection to examine this argument in detail.

16.6 Overcoming Alienation III: The *Geist* in the Machine

One general aspect that should be made explicit about the Always Already argument is that it is not straightforwardly deductive in form. In fact, I am not even sure if it *can* be presented in a simple deductive form. The reason for this, as I have already mentioned, is that what the argument aims at is trapping the *Kammerdiener* in a 'pragmatic' contradiction, i.e. a contradiction between what he says and his saying of it, the act and the content of the act. In this way, it is vaguely analogous to a *reductio* argument: no one can and/or ought to be in the discursive position of the *Kammerdiener* because the position is somehow self-contradictory or, as we shall see, self-defeating.

Brandom presents the Always Already argument as 'traversing the moments' of the four meta-meta-attitudes discussed above in Section 16.4. Summarised, the whole picture looks like this:

- (1) First, the judge acknowledges that he is adopting a *stance*, rather than simply acknowledging a *fact*.
- (2) Second, the judge acknowledges that the stance is a *recognitive* one.
- (3) So the judge acknowledges that which stance he adopts *produces* a community of a certain kind.
- (4) Next, the judge must acknowledge that acting and judging (acknowledging and attributing, deliberating and assessing) implicitly presuppose (are intelligible only in the context of) *edelmütig* recognitive stances.
- (5) Finally, the judge must explicitly adopt such a recognitive stance and institute an *edelmütig* recognitive community. (ST, 597)

The key moment here is number 4, which marks the act of 'forgiveness' (*Verzeihung*) of the confessing subject by the confessing judge. Forgiveness, in Brandom's

interpretation, means a kind of recollection in which the confessing subject's commitments are integrated in their subjective contingency into part of the retrospective narrative in which they revealed something of the correct content of the concept:

From the point of view of such a reconstructive recollection, though the decision might have been caused by contingent subjective attitudes and justificatorily irrelevant circumstances, *what* was so caused was an application that was both correct and expressively progressive. That is, it was just what was needed for us to find out more about the real content of the concept. The experience of incompatibility is exhibited in its capacity as the engine of the conceptual, cognitive, and practical *progress*, rather than in its capacity as the mark of error and failure. (ST, 602)

So, what the confessing judge effectively does in forgiving the confessing subject is to make a judgement to the effect that, despite the contingent basis on which the subject's applications were made, they managed to reveal something of the correct content of the concept, not merely subjective conceptions. But, of course, to *judge* that the application was (partially) correct does not mean that it was really correct; it only means that the judge accepts parts of the confessing subject's applications by showing their proper place in his recollective narrative. As such, nothing seems to stop the judge from holding onto his *niederträchtig* confession that acknowledges the causal contingency of his own forgiveness.

Let us backtrack to step number 3. At this point, the judge (assuming that he is not obstinately 'hard-hearted', i.e. refusing to confess his own subjectivity) confesses that his original *niederträchtig* evaluation of the subject is itself ridden with the contingency of his subjectivity; his attitude of judging was not solely subjunctively sensitive to the norm, which she acknowledged while making the original judgement. This step achieves a kind of mutual recognition in reciprocal *Niedertächtigkeit*, which can be understood in my simplified model as the claim that all the scorekeepers become self-aware Jesters and Naturalists. The judge's *niederträchtig* forgiveness then sounds something like this: 'Your application of the concept in this instance is correct: Lassie is correctly called a dog. However, the cause of your application was not a genuine semantic norm, but merely a collection of contingent circumstances conjoined with your attitudes. Likewise, my assessment of your application as correct is only an effect of myriad causal contextual forces which I do not fully understand'.

But this means, according to Brandom, that the scorekeepers cannot understand themselves as knowers and agents in the Kantian sense, which implies the ability to bind oneself with norms, for no Jester thinks that anyone can be genuinely bound by a norm (ST, 596).

We must now ask – why is reciprocal *Niedertüchtigkeit* problematic? Why would the next fourth step be necessary, in the sense of being obliged by reasons, for the Jester community to take? By giving up the commitment to genuine discursive normativity, why should the Jestlers, by that token, stop treating themselves and each other as knowers and agents in the full sense of the word? After all, *prima facie* there seems to be no reason why they could not draw the *modus tollens* conclusion that genuine discursive normativity is not necessary for discursive intentionality, i.e. being a knower and an agent.

I think that there are two ways to understand the crux of the Always Already argument at this seminal point. On the one hand, we can understand the argument as identifying an incompatibility between what the *Kammerdiener* says and his saying of it – the act and the content of the act, saying and doing. On the other hand, we can understand the argument as diagnosing the incompatibility of the *Kammerdiener* as taking place between his explicit normative attitude and his implicit normative status – an incompatibility between contents.

At first sight, it may appear that the latter reading is what Brandom has in mind:

What [the *Kammerdiener*] implicitly commits himself to by what he does is not what he explicitly acknowledges. *The contents of the status and of the attitude are incompatible with one another.* Their structure ensures that arguments of this form will be only as good as the understanding of conceptually articulated activity on which they are premised. (ST, 617, my italics)

Although Brandom clearly claims that the Always Already argument identifies an incompatibility between the *contents* to which the *Kammerdiener* is committed, I don't think this reading can be all that can be said for the argument, for in that case, it would be question-begging. Why is that, exactly? The reason is that while it is uncontroversial what the content of the *Kammerdiener's* explicit attitude is (namely, he denies that there are genuine norms that would be necessary to explain the application of concepts), what is precisely controversial is how to describe his

'implicit normative status' as a discursive being, or what he is 'in himself'. Obviously the *Kammerdiener* is not obliged to accept Brandom's definition of discursive beings, which include genuine discursive normativity; he is as free as anyone to choose his own definitions. So, what then is the argument that the *Kammerdiener* really is a discursive being in Brandom's normative sense, i.e. how is the content of his implicit status to be described? Where the Always Already argument is supposed to show that the *Kammerdiener* is implicitly a genuinely normative being *qua* a discursive one, it cannot presume that he is because that would make the argument circular.

I shall first look at what appears to be Brandom's main line of argument, i.e. that the incompatibility diagnostic of the *Kammerdiener* occurs between the contents of his explicit attitude and implicit status.

16.6.1 Reading 1 of the Always Already argument

As stated above, the first way to understand the Always Already argument diagnoses the incompatibility that the *Kammerdiener* is committed to between his explicit commitment to *Niederträchtigkeit* and his implicit status as a discursive being. Moreover, it is the contents of the attitude and the status which are supposed to lead to the pragmatic contradiction. As far as I can see, Brandom's main justification for this reading comes in the following passage:

We are now in a position to put meat on the bones of the fourth alternative [the cognitive-objective meta-meta-position]. [...] It acknowledges that the attitudes are cognitive ones, hence practical in the sense of *making* something to be so, not just *taking* it to be so. But it recaptures, at a higher level, versions of the objectivism and cognitivism of the first attitude. There is a kind of fact involved, which one would be ignoring if one adopted the *niederträchtig*, reductive attitude. That fact is the conceptual fact that determinate conceptual content and *Edelmütigkeit* in the form of confession and forgiveness are reciprocally sense-dependent concepts. Becoming explicitly aware of this fact is achieving the kind of self-consciousness characterized by *sittlich Vernunft* rather than alienated *Verstand*. Realizing it is realizing that in treating one's own thoughts and intentions as being determinately contentful, as binding one, making oneself representationally responsible to objective things in the sense that only certain ways the world could be would count as making one's beliefs true and one's intentions successful, is implicitly committing oneself to understanding oneself in terms of a community whose constitutive cognitive structure is that of reciprocal confession and forgiveness. Commitment to *Edelmütigkeit* is implicit in being a discursive being.

Alienation is having one's explicitly acknowledged commitments be incompatible with this structural commitment of consciousness and agency. (ST, 614–615)

The key claim here is the 'conceptual fact' that 'determinate conceptual content and *Edelmütigkeit* in the form of confession and forgiveness are reciprocally sense-dependent concepts'. In this subsection, my objection seeks to show that even if one accepted Brandom's account of both determinate conceptual contents and *Edelmütigkeit*, it does not follow that these concepts are reciprocally sense-dependent.

Moreover, it is important to observe that even if this objection fails, it will not follow from the claim that determinate conceptual content and *Edelmütigkeit* are reciprocally sense-dependent concepts that there are genuine discursive norms. For it is possible that these concepts are sense-dependent without being reference-dependent. In other words, even if it is true that no one can understand the concept of determinate conceptual content without understanding the concept of *Edelmütigkeit*, or being committed to adopting *edelmütig* attitudes, it is possible that there are, in fact, no genuine discursive norms to which one could refer. This is analogous to how one could not arguably understand the concept of hammer without the concept of nails without it, being the case that there could not be hammers without there being nails. However, in this work, I am not going to pursue this path of argumentation.

Onto the objection. To begin with, we must recall the broad outlines of Brandom's account about determinate conceptual contents, as discussed in Section 15. That account relies on three core ideas. The first is to explain 'of-intentionality', or Fregean reference, in terms of 'that-intentionality', or Fregean senses. Second, how this happens is by explaining the Fregean reference as an ideal sense playing a functional role in the judgement-forming system, such that whenever two (or more) commitments are found to be materially incompatible with each other, the subject will treat one as presenting a mere appearance and the other as presenting the reality by holding a certain auxiliary set of commitments fixed in the context. The third idea is that the judgement-forming mechanism can be said to represent something transcendental to the system (e.g. the stick) because the world itself is conceptually structured, so that when all goes well, the same propositional content appears in two modal forms, deontic and alethic, respectively. This is the process of experience that

explains how determinate conceptual contents are made possible by the process of finding them.

Now, Brandom claims that anyone who accepts this account about determinate conceptual contents is, by that token, implicitly committed to understanding oneself from the *edelmütig* meta-meta-perspective because these concepts are reciprocally sense-dependent, i.e. practically intelligible only in relation to one another. In other words, accepting the first commitment while rejecting the other leads to a material incompatibility. If true, it follows that there can really be no such thing as a *valid* '*niedertüchtig* confession', as I called it above, because every such confession would succumb to the incompatibility.

I think this diagnosis is incorrect. To begin with, recall that for the commitments *p* and *q* to be materially incompatible means that one cannot be entitled to both at the same time. Does this hold with the *niedertüchtig*, confessing judge? His first commitment is to the claim that a certain recollective, forgiving account of the confessing subject is correct; his second commitment is that in giving this forgiving recollection, what the judge (as anyone else) does is *not* explainable by reference to genuine discursive norms because there are no genuine discursive norms, only attitudes and causal circumstances. 'But', might Brandom say at this point, 'here is where the judge's commitments must be incompatible, for the content of the first commitment essentially includes the concept of (genuine) correctness, the existence of which the second commitment then goes on to deny. So, what the judge *does* (hold a certain forgiving recollective story as correct) is incompatible with what he *says* (denies that there are genuine norms that could determine the correctness of any such story)'.

But this is too quick. The crucial point is not *how* the judge performs his forgiving recollection, but *what* it is that he does in performing it. For the judge can agree that he can only *express* his endorsement of the forgiving recollection by using the discursive normative vocabulary, most importantly, the locution 'is correct'. This amounts to saying that one cannot generally purport to justify anything except by using a vocabulary apt for expressing justifications, i.e. the normative vocabulary. But it does not follow that, just because it is necessary to express the first

commitment by using normative vocabulary, the normative terms are necessary for *explaining* what is actually done by forgiving recollection.

Compare the case with the vocabulary of indexical expressions. There is widespread agreement nowadays that indexical expressions (e.g. 'I', 'you', 'him/her') are not synonymous with any non-indexical expressions, although it might well be true that it can be perfectly adequately explained what it is that one does by expressing sentences with indexical expressions without using them. This is essentially what Kaplan's concept of indexical 'character' is meant to do (Kaplan 1989).

In order to grasp the core of my reasoning for why the *niedertüchtig* confessing judge does not, in fact, commit an incompatibility, it is helpful to transform his allegedly incompatible commitments into the following, simplified claims:

1. Normative vocabulary expresses normative truths.
2. Normative truths do not explain the applications of concepts.

These claims are not materially incompatible; one can be entitled to both at the same time. The key point is that the *niederträchtig* judge can hold a minimalist account of normative truths, much like Kripkenstein's sceptic does, according to Kusch's reading as was explained in Section 7.1.1. By adopting reductive naturalism about discursive normativity, the *niederträchtig* judge need not hold that all normative claims are false. Rather, these claims can be used to express truths (the judge himself does that when he acknowledges a commitment to reductive naturalism), but truths only in a minimalist, deflated sense in which there is nothing more to the content of the truth predicate than the disquotational scheme. So, when the judge says that a certain forgiving recollective story is the *correct* thing to believe about the confessing subject, this is compatible with the judge's second claim that the use of the term 'correct' can be adequately explained without reference to genuine discursive norms.

Of course, Brandom has a contrasting account of what is, in fact, expressed by normative vocabulary, namely the *edelmütig* view, according to which what is expressed are genuine discursive commitments. But the crucial argument that I wanted to make has already run its course when we see that the *niederträchtig* judge does not commit an incompatibility when he refuses to endorse the *edelmütig* view.

In order to show the incompatibility, one should show that the judge cannot use normative vocabulary while being entitled to a theory, according to which normative truths do not explain the applications of concepts. But that is not what the Always Already argument shows; the Argument by itself does not engage at all with the entitlements, which the *niederträchtig* judge claims for the reductive view. Instead, the Always Argument seeks to strike earlier by showing that it is somehow incoherent to so much as try to defend such a view in practice.

It can be granted, however, that there is something odd about the *Kammerdiener's* claim that no one has, or can have, genuine reasons for applying concepts one way rather than another. For then it follows that the *Kammerdiener's* own application of concepts in endorsing the reductive view is without the force of genuine reasons in the sense that these reasons would explain why he applied his concepts as he did. I will consider this sort of objection at the end of the chapter, but for now, I want to press the observation that, even if this objection is cogent, it does not follow that the *Kammerdiener*, in holding onto a *niederträchtig* confession, commits an incompatibility. The reason, as I have already stated, is that it is one thing to *express* a normative claim and another to *explain* the applications of concepts with genuine (i.e. non-deflated) normative vocabulary.

16.6.2 Reading 2 of the Always Already Argument

I shall now move on to discuss what I take to be the second possible reading of the incompatibility diagnostic of the *Kammerdiener* that is supposed to drive the Always Already argument. To recapitulate, the first route tried to identify the incompatibility between contents to which the *Kammerdiener* is (necessarily) committed. This, I argued, fails. But perhaps there is a way to make sense of the diagnostic incompatibility by locating it between the *act itself* and its content, rather than between the contents of the *Kammerdiener's* explicit attitude and implicit status.

To begin with, we should take another look at the 'structural commitment' of scorekeepers to adopt the *Edelmütig* meta-meta-attitude (see the long quote at the beginning of the previous subsection). Insofar as commitments go, this one is special in at least three respects.

First, unlike ordinary practical (or theoretical) commitments, the structural commitment can only be undertaken implicitly. To undertake it explicitly (as opposed to simply explicitly recognising that it has already been undertaken) would require that one could express the intention to become a discursive being prior to being bound by the structural commitment, which is impossible precisely because the structural commitment is supposed to be a necessary condition for being a discursive being. In other words, the undertaking of the structural commitment can only be something that is done prior to the possibility (at least on the part of the doer) of saying what is being done: it can only be undertaken implicitly.

Second, it is not logically possible to rationally *reject* the structural, practical commitment to adopt the *Edelmütig* attitude. For a choice to be rational, according to Brandom, the agent must be bound by norms – at a minimum, the fundamental discursive norms. But since no subject is bound by (discursive) norms prior to undertaking the structural commitment, it is not possible to rationally reject it. It should be borne in mind, though, that the choice here is determined in the retrospective sense: it is not necessary (except perhaps in some causal sense) that any pre-discursive being becomes a discursive one, but once one develops discursive abilities, the structural commitment will already always have been undertaken.

Third, the structural commitment is unrejectable in the additional sense that once it has been undertaken, the subject cannot rationally relinquish it either. Of course, as Sellars says and Brandom is wont to repeat, one could always not speak at the cost of having nothing to say (see, e.g. Brandom 2008). However, that does not imply relinquishing the structural commitment so much as discursivity as such.⁹⁵ The reason why the structural commitment cannot be rationally rejected after it has been undertaken is the same as why it cannot be rationally rejected prior to being undertaken; because in both cases, the very act of explicitly expressing the rejection would simultaneously imply implicitly re-undertaking it. Since any discursive act is possible only supposing the implicit structural commitment, no discursive act can so much as purport to detach itself from it.

⁹⁵ Imagine there was a drug, which (for a period) deprived the user of all rational functions. I am sure there are a number of reasons why one could be entitled, in a given context, to take the drug. The point is, though, that there is no context in which one could rationally choose to relinquish the structural commitment in favour of some other, formally equally good commitment, which is the sense of relinquishing that matters.

At this juncture, it should be clear that the structural commitment is not really a 'commitment' in the ordinary sense, which implies the freedom of choice. That there is no freedom to choose seems incontestable insofar as choice ordinarily implies at least the logical option to refuse, which is not the case with the structural commitment as we just saw. Moreover, it is plain that as a 'condition of possibility' for having discursive abilities, the structural commitment is different from any causal or physical conditions that are, in a sense, necessary for having those abilities. Rather, the structural commitment depicted above is close to some kind of a fate that is imposed on the subject as soon as she becomes a rational subject.

Another way to describe the undertaking of the structural commitment to *Edelmütigkeit* is that it is a 'pure act', i.e. an act without propositional, conceptual content, or where this kind of content is irrelevant to its entitlement. As we just saw, the structural commitment cannot be undertaken in the way ordinary commitments can be because it is a condition of possibility for there to be any commitments at all; it cannot be part of the ordinary normative-inferential network of commitments and entitlements. Moreover, it is because undertaking the structural commitment is *pure* that its binding force must essentially be *retroactive*. If at t_0 , the subject counts as non-discursive, and at a later moment t_2 , she counts as discursive, then there must be a moment t_1 between t_0 and t_2 , at which she becomes discursive. But what the retroactivity of the structural commitment means is that t_1 will not occur linearly between t_0 and t_2 , but rather *will have* occurred once t_2 has. The continuity from t_0 to t_2 is non-linear and retroactive, but it is continuity, nonetheless. It should be emphasised that at stake is not retroactive *causal* change, which has its own problems, for the structural commitment is pure and has no content; it changes nothing at the level of contents undertaken because it is a condition of possibility for there to be any content at all. The retroactive effect is clearly claimed by Brandom when he discusses the consequences of the forgiveness-qua-recollection account:

But invoking the practical recollective work that is the recovery of an intention as a concept-application that unifies the purposive and consequential aspects of actions points to the way in which forgiveness on the practical side can be not only *retrospective*, in reconstruing what is taken to be objective content of the concept toward which a practical attitude is adopted in endorsing a purpose, but also *retroactive*. (ST, 624)

Something I have done should not be treated as an error or a crime, as the hard-hearted *niederträchtig* judge does, because it is not yet settled *what* I have done. (ST, 625)

To get a feeling as to how the second way of reading the diagnostic incompatibility is supposed to go, let us make a comparison to how Jaakko Hintikka analyses the assertion 'I don't exist'. This sentence is false every time it is asserted by anyone, while its negation is always true when asserted by anyone. Yet, the claim made by the assertion does not express a necessary falsehood, nor its negation a necessary truth. Neither does the assertion's content conceal a contradiction, for if we replace the 'I' with a proper name, 'N.N.', the result is not a self-contradiction even when uttered by N.N.⁹⁶ How can this be? According to Hintikka, the reason is that by making the assertion 'I don't exist', one by that token *performatively demonstrates* the falsity of the assertion, just like by asserting (or thinking) that 'I exist', one demonstrates the truth of the assertion either to oneself or to the audience. The very *act* of thinking or asserting is what carries the weight of demonstration, not the *content* as such. 'I don't exist' represents perhaps the purest form of pragmatic contradiction precisely because it is most easily seen as *self-defeating* (immediately demonstrates itself to be false, albeit perhaps fallibly so) to and by anyone who understands the sentence's meaning. (Hintikka 1962, 15–16)

This is then how Hintikka analyses the assertion 'I don't exist'; it is not self-contradictory but *self-defeating* in the sense that no one can *persuade* anyone, including oneself, to believe the proposition they express by asserting 'I don't exist'. Every time one asserts that 'I don't exist', the very act of assertion 'defeats' what is being asserted. Perhaps something similar applies to the *Kammerdiener*?

To begin with, showing self-defeatingness in the *Kammerdiener's* case is much more complicated – or at the very least, less intuitive – than it is in the case of 'I don't exist'. Returning to step number three of the traversing story above, if we imagine ourselves in the place of the Jesters, how exactly would self-defeatingness work here? The key point that seems to recur is: how is it to be decided what the status of the

⁹⁶ It is not clear whether even the sentence 'N.N. says that N.N. does not exist' is self-contradictory. For arguably, that is something which Hamlet could have said; would it follow that Hamlet existed? This difficult question can be sidelined here.

scorekeepers is, to wit, is it 'normative' or 'merely' natural? This is not intuitive in the sense in which the demonstration of one's existence is.

Take another example of a typical pragmatic contradiction: 'I am a very humble person.' Is that something a truly humble person would say? Intuitively not, although context does matter, of course. Notably, there seems to be very little in the way of 'theoretical content' required to make sense of the pragmatic contradiction. All that is needed is a practical grasp of the meaning of 'humble' and a sense of the illocutionary force of the assertion (i.e. that it is braggy). Of course, it is possible that the speaker in this case has a different 'theory' or practical grasp of what it means to be humble than his audience does, or then the speaker interprets the illocutionary force differently than his audience does. The pragmatic contradiction depends on both of these factors; in that sense, it is not only about the act and its content.

Try to imagine the *Kammerdiener's* case in a similar spirit. Someone says: 'There are no genuine discursive norms'. Another responds: 'Well, how can you speak then?' This response comes across as question-begging because it seems to assume more than the ordinary meaning of 'genuine discursive norms' (supposing that the expression does have some ordinary, intuitively graspable meaning). So, some theoretical commitments are needed to make sense of the response. But then we have arguably moved beyond the paradigmatic cases of pragmatic contradiction considered in the second reading's sense; the alleged incompatibility does not simply spark between the speaker's act and the content of the act. But since we just saw this is not strictly speaking the case with 'humble' either, perhaps this is not a problem? In both cases, what is needed is some 'theory' (or practical grasp) about the meaning of 'humble' (or for that matter, 'to exist'), as well as an interpretation of the illocutionary force of the speech act.

I think there are two ways to respond to these observations. One is to agree that the case of 'humble' and that of the *Kammerdiener* are similar, but insist that the latter is clearly more 'theoretically loaded' than the former. You need to have more 'theory' to respond to the *Kammerdiener* in the way that Brandom does, and this means the issue is no longer about a pragmatic contradiction in the paradigmatic sense of 'I don't exist'. This is a difficult case to argue for conclusively, for it is rich with nuance and presuppositions.

The second way to respond sees that there is, after all, no strict separation between two readings of the pragmatic contradiction diagnostic of the *Kammerdiener*, for both have to include facts about the contents of his commitments in some fashion. It is never *just* the act and its content that are contradictory or incompatible; it is a matter of degree how much content or 'theory' is needed for a pragmatic incompatibility. This way seems, to me, a more natural conclusion to hold, or at least easier.⁹⁷

In either case, I think it can be concluded that the Always Already argument does not work as Brandom intends. That is to say, the four moments discussed in Section 16.6 need not be 'traversed' in the manner that he presents; it is rationally possible to remain at the third level and to refuse endorsing the *edelmütig* view. What makes this possible is the distinction between the expressive and explanatory uses of 'semantic correctness' where the expressive use is understood in the deflationary sense and denied a robust explanatory role. This move is not something which the Always Already argument can defeat by itself because it does not show that discursive normative truths could not be deflated truths, i.e. truths that are exhausted by the disquotational scheme.

What remains is to answer the objection mentioned at the end of the previous subsection. If reductive naturalism about discursive normativity is true, no application of concepts is explained by genuine norms or reasons. Does it not follow that our use of words is therefore rationally groundless, that there is no reason a word should be used one way rather than another? Answering this worry is one of the main tasks in the last section of the chapter.

⁹⁷ Conclusions along these lines are also found in Errázuriz (2014), who discusses the pragmatic contradiction at a more general level.

17 BACK TO KRIPKENSTEIN

My purpose in this section is twofold. The first goal is to argue that a certain kind of reductive naturalist metasemantic theory can, and should, accept three of Brandom's four epistemic criteria of adequacy discussed in Section 15.1. In particular, I think Devitt's work, as cursorily examined in Section 7.3.2 of Chapter I, is naturally compatible with the Genuine Knowledge Condition, the Intelligibility of Error Condition, and the Mode of Presentation Condition. However, the second goal is to argue that although the naturalist cannot account for the Rational Constraint Condition, this is not something she is required to do, for RCC is not a genuine criterion of adequacy on semantic theories. Rejecting this criterion effectively means rejecting the ideas that (i) meanings would provide *genuine* reasons for application and that (ii) meaningful use must be justified use (in Brandom's sense of justification, at any rate). I shall begin by arguing for the second idea first.

In the Conclusion chapter of *ST*, Brandom comes to review his recollective, magnanimous (i.e. *edelmütig*) story in the light of Kripkenstein's sceptical challenge, laid out in three steps:

(1) The observation that sets the stage for Kripke's Wittgenstein's argument is that the determinate contentfulness of normative attitudes is intelligible only insofar as those attitudes are understood as responsible to norms—that is, only insofar as normative statuses, in the form of the commitments one undertakes by believing or intending, are authoritative with respect to assessment of the correctness or success of the attitudes in question.

(2) Kripke's Wittgenstein's second claim is that only the *use* of the expression can confer that content on it. *All* there is to determine the content of the concept applied, and so the norm that governs applications of it (in the sense of providing a standard of normative assessments of correctness for those applications), is the way it has been applied, the attitudes that have in fact been adopted.

(3) The third move in the argument is then the claim that there is no way to explain how any course of past actual applications of a concept can determine a normative standard for assessing the correctness of novel possible future uses. (*ST*, 650–652)

As a summarisation of the sceptical challenge, the major problem, according to Brandom, is the following:

That as a matter of contingent fact we can catch on to an extension of prior practice, that we can all learn to agree in practice about ‘how to go on’ to apply it in new cases, is not a *reason* to go on that way. ‘We just talk that way,’ is an observation about our practices that is not at the right level to serve as a justification for claims about the world to the effect that one thing is to be found in another—any more than offering a historical explanation of why we use the sign designs ‘dog’ (written and spoken) to refer to dogs, instead of some other arbitrary marks and noises, justifies the claim that Lassie is a dog. (ST, 657)

According to Brandom’s reading of Kripkenstein, the initial, unchallenged thought is that the meanings of words provide reasons for applying them one way rather than another, and that this feature of meanings as standards of justification on use is somehow essential to the very concept of meaning (determinate contentfulness). One important dimension of Kripke’s exposition examines the kinds of criteria that go into such reasons, as well as their possible origins. The key issue that I traced in Chapter I concerns the point that is also of fundamental importance to Brandom: in what sense can the reasons be ‘determinate’? Another related issue concerns normativity: if meanings serve as reasons for the application of words, in what sense are they ‘normative’? Put alternatively, in what sense can (or must) meaning ‘justify’ use?

The ‘paradox’, which results from the sceptical challenge, is rephrased by Brandom in terms of alienation. I think that for this purpose, it is not even necessary to accept the premise that only facts about actual use can determine what meaning a given word comes to have; it suffices to see that, insofar as the problem of finitude goes through (as I have argued it does), absolute determinacy is false for finite speakers. Meanings cannot determine an infinite number of applications; hence, it is always metaphysically an open question whether a novel application counts as going on the same way as previous ones. Alienation ensues, therefore, as a discrepancy between the metasemantic fact that meanings are not absolutely determinate and the premise that meanings provide normative reasons for application. If there is no metasemantic fact, which would determine what would count as using a word in the *same* way as before, it seems that there can be no normative reason to apply a word one way rather than another. To Brandom, that is unacceptable since it would mean that a determinately meaningful application becomes impossible, for it would be just as

contingently arbitrary as the fact that 'dog' is the sound form referring to dogs. Hence, the reductive naturalist metasemantic account, if true, would, according to Brandom, make it impossible for the subjects to understand their usage as genuinely normative (i.e. justified), which is the hallmark of alienation.

It should be emphasised that what precipitates alienation is precisely the normativity problem (understood as an issue of justification) and not the finitude problem. In fact, the *Vernunft* meta-metatheory of concepts that Brandom espouses is more or less in line with the rejection of absolute determinacy, at least in the *Verstand* sense. Meanings do not by themselves extend to cover new applications; their determination is a function of their application, something that must be (in part) made and not merely found, which captures an important aspect of meaning finitism. Brandom's guiding idea is that for the process of extending old uses to cover new applications to work, it is necessary that the extensions are justified by reasons in a relatively strong sense. In particular, the normativity involved in such reasons must amount to something more than a uniform intersubjective 'agreement', which would be a fact just about the attitudes. Indeed, the normativity must be strong enough to exceed what subjects may be able to achieve: the 'ought' overrides the 'can' (BSD, 191). Only in this sense can the process of semantic determination be properly called 'rational'; only then can the fourth epistemic criterion of adequacy (the Rational Constraint Condition) be satisfied.

It will be useful to look a little closer at what kind of semantic demands on epistemic theory the RCC really implies. Ordinarily, when we want justifications for calling Lassie a dog, the reply consists of something along the lines of 'She barks, chases cats, has a certain genetic make-up, etc.'. These things are evidence of the property of being a dog. But the important thing is that Kripkenstein's sceptic does not directly challenge the cogency of *that* kind of evidence, no more than she challenges the mathematical result that $58 + 67 = 125$. The paradox is meant to locate at the metasemantic level, where the question is how the word 'dog' came to have its ordinary use-justifying meaning (i.e. 'chases cats, barks, has a certain genetic make-up, etc.'). Since the only base explanans to this question can draw from actual use (or in general, from some other finite set), the argument goes: the word 'dog' cannot have a determinate meaning, for the merely descriptive facts about use are not 'fit' to be *cited as reasons for extending* the past use in the *same* way to the novel instances

of application. The rationale behind RCC then comes down to this: the non-normatively described fact that in the past, we have called animals like Lassie 'dogs' is not a reason for (not) calling Lassie a 'dog', i.e. to extend the use one way rather than another. Hence, the use, considered as only something that has happened and continues to happen, cannot be how words come to have their meanings *if meanings must essentially justify usage* (or if meaningful use must essentially consist of semantically justified use). But if use is, by premise (2), the only thing that can confer meanings on expressions, then the conclusion must be that meaningful language is impossible.

In the conclusions of ST, Brandom also expounds on his reading of the sceptical solution that he thinks Wittgenstein himself endorsed. Borrowing terms from Heidegger, the semantic alienation described above results in radical semantic *Geworfenheit*, being thrown into a world of contingently existing norms that structure our linguistic behaviour without our ability to justify them. The 'sceptical solution' to *Geworfenheit* is not to find new metaphysics to execute the justifying, but rather to learn to live with it in the spirit of semantic *Gelassenheit*:

Any account of discursive normativity that treats the fact of our semantic *Geworfenheit* as undercutting the legitimacy of those norms (that is, any alienating account) is to be rejected as incorporating an evidently mistaken metaphysics of normativity. The proper response to this realization, Wittgenstein thinks, is not to construct some alternative positive metaphysical story, but simply to acknowledge and embrace discursive contingency and semantic *Geworfenheit*. We might call this recommended therapeutic meta-attitude 'semantic *Gelassenheit*,' to continue the Heideggerian metaphor. [...] Basically, it recommends that we just get used to our *Geworfenheit*, rejecting theories according to which it is alienating, without adopting others in their stead. (ST, 658–659)

However, according to Brandom, semantic *Gelassenheit* must not only be rejected; it must also be understood as falling prey to the Always Already argument and pragmatic contradiction, just as other forms of alienation. For the only way in which that argument works is by showing that the *edelmütig* metanormative attitude is a necessary condition for being a speaking, thinking being. The reason why the commitment to *Edelmütigkeit* must be *fated* is that, were it simply a contingent, subjective stance among others, it would lose its ability to explain genuine normative force.

Let us now return to the Lassie example. Here, we have two different kinds of questions:

- (1) The question whether Lassie is a dog.
- (2) The question whether 'is a dog' applies correctly to Lassie.

On the surface, it seems that there must be some kind of a connection between these two questions. Indeed, I think there is; but the way how the connection works is understood very differently by Brandom and by the causal-historical account.

Let us take another look at how Brandom understands the connection. In order for the ground-level use of empirical concepts, like the predicate 'is a dog', to have determined meanings, he thinks, there must be an answer to question number (2). Moreover, the answer must be such that whatever determines that 'is a dog' applies correctly to Lassie can serve as a genuine reason for the ground level applications, as represented by question (1). This is why merely causal-attitudinal facts about the history of the predicate expression 'is a dog' are insufficient to give an answer to (2) – because such facts cannot be cited as reasons to answer (1).

How does the causal-historical picture understand the case? We can start by looking at how subjects, in fact, go on to answer questions like (1). Typically, the speakers refer to some identifying criteria, which are more or less readily accessible for judgement, e.g. facts about how Lassie looks, behaves, and what her genetic make-up and origins are. If some of these criteria turn out to create a conflict in this instance, it might be more or less clear which are most important, and which can be ignored. However, if all goes well, the speakers will eventually reach a decision one way or another; then they have reached the (defeasible) reasons they can cite to support their application.

However, according to the causal-historical picture, it is *not* the descriptive criteria that the speakers indeed rely on to make their case that determine the reference (i.e. correct application) of 'is a dog' in this instance. For it could turn out that most or even all of these criteria are, in fact, false about the entities called 'dogs'. Our best evidence as to the criteria of doghood could all be massively mistaken if, say, dogs turned out to be Martian spy-robots. But according to the causal-historical account,

the speakers would still have been referring to dogs by 'dog': the descriptive criteria the speakers have actually used could not have been what determined the reference here.

What determines the reference of 'is a dog', according to the causal-historical account, is, of course, the D-chains of borrowing and grounding that circulate in the historical community. Importantly, these chains can determine reference irrespectively of the fact of whether the speakers know that they do, or whether they indeed use such chains as descriptive criterial evidence of doghood. In fact, in most cases, they do not, the reason being that it would be tautologous to do so unless the D-chains themselves were vindicated by other means. Brandom is certainly right that to say that 'Lassie is a dog because she is the kind of entity to which our speech community has previously referred to by the expression "is a dog"' is not much of an argument to call Lassie a dog.

The important point, now, is this. Although according to the causal-historical account, the descriptive criteria that the speakers have to identify dogs are not *immediately* what ground the reference of 'is a dog', these criteria and the reasons they stand for play a *mediate* role in determining the reference. For these criteria are both the reasons *and* causes of the 'is a dog' applications; and it is these applications which (multiply) ground the reference. My explicit reasons for calling Lassie a dog is that she barks, chases cats, etc. And these facts are also part of the causal story that explains why I applied the expression as I did. For if Lassie did not meet most or any of the criteria, it is unlikely that I would call her a dog. Still, for all I know, she could *be* a dog.

We can now see that the causal-historical account can, in fact, agree with the claim that the correct applicability of expressions is in part determined by the *reasons* speakers explicitly cite to support their applications (or at any rate, reasons which the speakers *could* cite as justifications, even if they do not bother to do so in every particular case). The way how the reasons (i.e. descriptive criteria as cited by the speakers) in part determine the reference is by influencing the actual applications of the speakers. The difference to Brandom is that he wants something more out of the metasemantic theory than this mundane sense of reasons: he wants the reasons themselves to be justified in a further, higher sense. Something must justify the

reasons speakers, in fact, use to justify their applications, and this something is the 'concept itself', i.e. the genuine norms, which, according to him, constitute propositional contents and which are instituted by reciprocal, forgiving recollection.

The major point that I want to make here is that, unlike Brandom claims, an account that does not include genuine discursive norms can still make sense of the idea that applications are sensitive to normative reasons. If it does turn out that dogs are indeed Martian spy-robots or whatever, that is a *good* reason to change our descriptive criteria, and thus the ordinary meaning, of 'is a dog'. No further justification is required for this reason itself, even if the change is driven by merely causal and contingent facts. Brandom might insist that this is not a case of applications being sensitive to 'genuine' reasons, which would rely on some retroactive recollective forgiving story. But is that something that the causal-historical account needs to bother with, really? In other words, is 'genuine discursive normativity' kind of a pretheoretical standard on metasemantic theories, comparable, e.g. to the compositionality of meaning?

I think that what drives Brandom to look for the 'extra normativity' of applications is that he wants to give a straight answer, not only to Kripkenstein's sceptical challenge, but to a version of Agrippa's trilemma. We can see how the trilemma is analogous to the sceptical challenge if we imagine the sceptic challenging the speakers to provide justifications for the criteria they use to identify dogs, then justifications for these justifications. Since the use of descriptive criteria effectively rests on the applications of other expressions, which in turn rest their own justifications on further descriptive criteria, the end result must either be (a) an infinite regress, (b) circularity, or (c) primitivism of some kind. Since Brandom seems to think that the trilemma is valid, and that all the options are faulty, the only way to resolve it is by telling a story of semantic justification in a way where the ordinary, descriptive criteria are justified by something more than further ordinary, descriptive criteria. That something more is the account of recollective forgiveness and recognition. The form of this solution to the trilemma is that there is a presupposition necessary for anyone to be in a position to pose the trilemma, which suffices to refute it as well, namely the presupposition that discursive beings are by their nature already always committed to adopt *edelmütig*, forgiving attitudes towards

themselves and each other – and by that token, effectively to pull each other out of the mire.

In the previous section, I argued that the account of recollective forgiveness does not succeed in refuting the *Kammerdiener*. Here is my argument for why we *need not* refute the *Kammerdiener* in this sense, but rather can learn to walk in his shoes. The answer consists of two premises, which we have already accounted for: the rejection of absolute determinacy and the endorsement of the causal-historical account of reference. The first premise grants to the sceptic that in general, there is no 'superlative fact', formulable in terms of Fregean determinateness, that would settle, e.g. the absolutely correct applicability of 'is a dog'. The second premise follows this up by understanding 'correct applicability' in terms of the causal-historical reference that is determined temporally and historically by the D-chains of grounding and borrowing circulating in a speech community. Although the sceptic can, in principle, always challenge the justifications for the descriptive criteria we actually use to identify dogs, this is not the same as challenging the very idea of the determined correct applicability of 'is a dog', which is not *immediately* determined by the descriptive criteria. The correct applicability, i.e. reference, of 'is a dog' is an empirical matter logically (though not causally) independent of the descriptive criteria. Thus, we can retain the idea of determinate conceptual content while granting the sceptic the first premise.

What about semantic justification and the looming trilemma? My idea is that once we accept the aforementioned solution to how the conceptual contents of our applications can be determinate, we thereby solve the justification problem as well. For once a predicate expression has been grounded in a referent, we have established a standard for a correct application that is logically independent of the criteria we use to identify tokens of the referent. Challenging the criteria will not challenge the standard itself. So, even if there is no way out of the trilemma in the sense of a global regress stopper that would justify *all* the ground-level reasons we have for applying expressions, this does not mean we cannot use the criteria we have in every particular case where the justification of the expressions comes up, while also retaining a determinate standard for applications. In other words, it turns out that semantic justification in the form of forgiving recollection is not necessary for determinate conceptual contents.

Of course, to talk about 'determinate conceptual contents' means something else in the context of causal-historical meaning finitism than it does in the context of absolute determinacy or the normativist *Vernunft* view. As was mentioned, Brandom wants to hold on to Fregean determinateness as a regulative ideal for concepts and their application. But he does not give arguments as to why it would be *necessary* to retain Fregean determinateness even in the *Vernunft* sense. Insofar as meaning finitism is internally coherent, it is possible to define 'determinate conceptual content' differently rather than give up the very idea. Differently how, exactly? Admittedly, that is a promise that further work must cash out.

17.1 The Causal-Historical Account and the Four Semantic Criteria for Epistemic Theories

Next, I shall focus on how the causal-historical account of meanings can both fulfil and benefit from three of Brandom's four semantic criteria of adequacy for epistemic theories, introduced in Section 15.1. To begin with, the core causalist picture of meanings consists of the following three ideas:

- (i.) Rejection of absolute (or Fregean) determinacy and endorsement of a historical, finitist account of meanings.
- (ii.) Knowledge of meanings is empirical, hence non-privileged. (Fulfil IEC.)
- (iii.) Meaning *qua* sense is an aspect of the causal-historical reference. This explains both how our cognitive grasp of meanings is empirical and why senses essentially determine their referents, or why a given token sense is *about* a given token referent. (Fulfil the GKC and MPC.)⁹⁸

A key thematic idea of the causal-historical account of meaning, as I understand it, is that *the purpose of a theory of reference is not to explain the truth values or truth conditions of*

⁹⁸ To be precise, all these ideas should be tensed in the present perfect: A given sense *has* until now been about a given referent.

*meaning sentences but rather certain truths about our classificatory linguistic practices.*⁹⁹ This is because the purpose of the theory of reference is not to justify those practices in the metalinguistic (or metasemantic) sense, only to describe and explain them in the semantic sense. As argued above, the justification of ordinary empirical concepts does not need a global regress-stopper, and hence fulfilling the RCC is not a genuine criterion of adequacy on metasemantic theories.

Next, I shall briefly show how the causal-historical account of meaning can meet the first three semantic criteria of adequacy for epistemic theories. This discussion concerns the points (ii) and (iii) mostly, for as I argued in Chapter I, abandoning absolute determinacy is necessary for any metasemantic theory.

First, how does the causal-historical account's claim that the knowledge of meanings is empirical meet the Intelligibility of Error Condition? As we recall, the core feature of so-called two-stage representational theories was that there are two kinds of knowledge-that: one about representings, and the other about representeds. The challenge posed by IEC is that for the knowledge of representeds to be genuine (and for the GKC to be fulfilled), errors must be possible on both the side of representings and representeds. Moreover, the knowledge of representings cannot be representational itself, for that would lead to a regress, so the error must be non-representational in kind. Devitt's solution meets this condition essentially because he understands the knowledge of meanings to be practical knowledge-how, not knowledge-that (Devitt 2001, 481). Being competent with a word like 'dog' does not entail that one has knowledge-that about the meaning, but only that one is able to use it in various contexts. But since the practical knowledge is not representational, as any skill in general is not, there is no risk of regress.

⁹⁹ Importantly, this is not how Devitt himself understands the aim of the account; according to him, one purpose of a theory of reference is to in part explain how sentences come to have truth conditions (Devitt and Sterelny 1999, 19). The reason why I think this should not be an aim of the theory of reference is that meaning finitism is incompatible with the idea that sentences would have determinate, unique truth conditions. However, it should be noted that what Devitt has in mind is a conception of truth conditions in terms of absolute determinacy. Is it possible to understand truth conditions without this commitment, i.e. in a temporal sense? While interesting and worthy of further study, in this context, I will work on the assumption that absolute determinacy cannot be detached from the very idea of truth conditions.

But what about our theoretical knowledge of meanings themselves, i.e. not the phenomenon competence but the ability of meanings to determine referents, their being modes of reference? Surely, this knowledge must then be propositional knowledge-that and representational? Indeed, it is, but there is no risk of regress, for our theoretical knowledge of meanings themselves is knowledge about worldly objects, i.e. meanings as causal-historical chains, and such knowledge is not necessary for being competent with a word. This is where the Mode of Presentation Condition is fulfilled. What is perhaps most shocking about Devitt's shocking idea is that meanings themselves are claimed to be causal-historical chains of reference, our knowledge of which, when available, is empirical. The reason why a token of a word, say the name 'Kripke', is about Kripke and not Kripnam is that the meaning of the word has (until now, at any rate) determined Kripke and not Kripnam as its referent.¹⁰⁰ How is this possible? The reason is that the meaning as a causal-historical chain is just the historical chain of tokenings of the name 'Kripke' joined with intentions to refer to whomever the person from whom the name was borrowed referred with it. How do we know that this is really the case? By the mundane empirical means in which we learn what person or object a given name denotes, e.g. being shown the person or a picture of him. The 'aboutness' of meaning in this sense ultimately amounts to the empirical fact that the D-chains of a name in a given community have historically, in various ways, causally ended in a given person or object. It goes without saying that in this view, our knowledge of meanings is much less readily accessible than descriptivist theories take it to be.

Naturally, having said this much, nothing much has yet been said to turn anyone's mind who was not already a believer in the causal-historical approach. The abundance of major, difficult questions that remain for future research to solve include details about the causal-historical chains, the role of intentions, whether the account can be extended from the paradigmatic case of proper names to other expression types, etc. I cannot hope to resolve all these and other problems in this work, although others more qualified have already done a great deal to that benefit. That being said, the purpose of this section was not to argue for the causal-historical account as such, only to show that it can indeed fulfil, at least broadly, three of the

¹⁰⁰ According to my supervisor Panu, who is well-versed in Kripkean literature, it is a common exegetical mistake to bundle Kripke's and Putnam's views about externalism and other related topics together into some kind of 'Kripke-Putnam' orthodoxy. Perhaps Colin McGinn's example is not so far-fetched as all that!

four semantic criteria for epistemic theories, which Brandom posits in ST. And as argued in the previous subsection, fulfilling the fourth criterion is not something it needs to do to begin with.

18 CONCLUSIONS

The most important claim that this work advances is that there is no straight solution to Kripkenstein's problem of finitude. While I could not prove the claim, I added to the reasons provided by Kripke and Kusch to take the problem of finitude – and its insolubility – seriously.

From this basic, tentative conclusion, I drew three follow-up claims:

- (1.) Ascriptions of meaning are not meaningful in virtue of expressing propositions with truth conditions. (I.1.4)
- (2.) Meanings, facts grounding meaning facts, and meaning ascriptions must be determined finitely. (I.6.2.1)
- (3.) Meanings, facts grounding meaning facts, and meaning ascriptions must be determined temporally. (I.6.2)

These claims may seem to entail an anti-realist metasemantic position. Roughly, this would amount to saying that meaning ascriptions do not count as descriptions (other than in a minimal, deflated sense as explained in Chapter I, 6.1.1), and that there are no facts which ground meaning facts, indeed that meanings are not factual at all except in a deflated sense. However, instead, I sought to show that the causal-historical theory of meanings, as developed by Devitt, can provide a reasonable realist theory of meanings, assuming that it can accommodate the rejection of absolute determinacy. I offered some tentative ways in which this could be done with the help of Gómez-Torrente's theory of reference fixing.

This much was achieved in Chapter I. In Chapters II and III, I examined at length Brandom's normativist, pragmatist, inferentialist, and expressivist theory of language and meaning. My main goal here was to read Brandom's ambitious, grand synthesis between phenomenalism about norms and normative phenomenalism as an

extended solution to the problem of finitude. The key here was the concept of genuine discursive normativity, which Brandom claims can be implicitly instituted in the scorekeeping practice. If successful, the account would provide a way to solve the problem of finitude while retaining a commitment to (something like) absolute determinacy in the *Vernunft* sense. I discerned two distinct strategies, one in MIE and another in ST, for reaching this conclusion.

Briefly, the strategy of MIE to answer the problem of finitude by answering the normativity problem goes through the concept of semantic objectivity. If the norms instituted by the scorekeeping practice can confer objective semantic commitments in the sense of incorporating worldly objects, the thought goes: the genuineness of implicit norms itself will have been vindicated. In the terminology I adopted, the pragmatic objectivity of norms is reached via the semantic objectivity of propositional contents.

However, although I showed (11.5-11.6) how Brandom is able, in principle, to arrive at semantic objectivity in MIE (namely, by including the metaphysical claim of conceptual realism), I also argued (12.5) that semantic objectivity is not sufficient to establish pragmatic objectivity. As far as MIE goes, the problem of reason's sovereignty, as I called it, remains standing (12.6), which then exposes the whole project to the anti-transcendentalist arguments made by Turner (12.8). This much was covered in Chapter II.

In Chapter III, I first construed Brandom's new argument for how to defeat Kripkenstein's sceptic and the *Kammerdiener* both. The form of the Always Already argument is to show how genuine discursive normativity can be implicitly instituted by mutual recognition by showing (a) that mutual recognition can be *de facto* symmetrical (and *de jure* transitive) and (b) that every scorekeeper is already always structurally committed to recognise each other's discursive authority. Second, I argued that the Always Already argument does not succeed in defeating Kripkenstein's sceptic or the *Kammerdiener*. The main problem is in showing how exactly the pragmatic contradiction is supposed to rationally oblige the sceptic to recognise genuine norms and discursive authority. In order to show this, more needs to be said in support of the Always Already argument and the pragmatic contradiction in particular.

REFERENCES

- Anscombe, E. (1985). 'Review: Wittgenstein on Rules and Private Language.' *Ethics*, Vol. 95, No. 2, 342-352.
- Armstrong, D. (1978). *Universals and Scientific Realism*. Cambridge University Press.
- Baker, G.P. and Hacker, P.M.S. (1984). 'On Misunderstanding Wittgenstein: Kripke's Private Language Argument.' *Synthese*, Vol. 58, No. 3, 407-450.
- Bayern, A.M.P.v., Danel, S., Auersperg, A.M.I. *et al.* (2018). 'Compound tool construction by New Caledonian crows.' *Scientific Reports*, **8**, 15676.
- Belnap, N. (1962). 'Tonk, Plonk and Plink.' *Analysis*, 22, 130-134.
- Bertram, G.W. (2020). 'Where Is the Conflict in Brandom's Theory of Recognition (and Why Should There Be Any)?' Published in *Reading Brandom: On A Spirit of Trust*. (ed.) Bouché, G. New York, Routledge.
- Blackburn, S. (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Clarendon Press, Oxford.
- Bloor, D. (1997). *Wittgenstein, Rules and Institutions*. London, Routledge.
- Boghossian, P. (1989): 'The Rule-Following Considerations.' *Mind, New Series*, vol.98, no.392, 507-549.
- Boghossian, P. (2012). 'Blind Rule-Following.' Published in *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*. Annalisa Coliva (Ed.) Oxford University Press, UK.
- Brandom, R. (2021). 'Replies to Honneth, McDowell, Pippin, and Stern.' *Philosophy and Phenomenological Research*, 103, 741–760.
- Brandom, R. (2019/ST). *A Spirit of Trust: A Reading of Hegel's Phenomenology*. Cambridge MA, Harvard University Press.
- Brandom, R. (2019). 'Some Strands of Wittgenstein's Normative Pragmatism, and Some Strains of his Semantic Nihilism.' *Disputatio. Philosophical Research Bulletin* Vol. 8, No. 9, 0-00.
- Brandom, R. (2011/PP). *Perspectives on Pragmatism: Classical, Recent, and Contemporary*. Cambridge MA, Harvard University Press.
- Brandom R. (2010a). 'Reply to Sebastian Rödl's 'Brandom's Theory of the Mind.'" Published in *Reading Brandom: On Making It Explicit*. (ed.) Weiss, B. and Wanderer, J. Taylor and Francis Group, ProQuest Ebook Central.

- Brandom R. (2010b). 'Reply to Daniel Dennett's "The Evolution of "Why?"' Published in *Reading Brandom: On Making It Explicit*. (ed.) Weiss, B. and Wanderer, J. Taylor and Francis Group, ProQuest Ebook Central.
- Brandom, R. (2009/RP). *Reason in Philosophy: Animating Ideas*. Cambridge MA, Harvard University Press.
- Brandom, R. (2008/BSD). *Between Saying and Doing*. New York, Oxford University Press.
- Brandom, R. (2008). 'Reply to "Are Fundamental Discursive Norms Objective?"' Published in *Robert Brandom: Analytic Pragmatist*. (ed.) Prien, B, Schweikard, D. Berlin, De Gruyter.
- Brandom, R. (2005). 'Responses.' *Pragmatics & Cognition* 13:1, 227–249.
- Brandom, R. (2002/TMD). *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge MA, Harvard University Press.
- Brandom, R. (2001/AR). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge MA, Harvard University Press.
- Brandom, R. (2001). 'Modality, Normativity, and Intentionality.' *Philosophy and Phenomenological Research*, Vol. LXIII, No. 3, 587-609.
- Brandom, R. (1994/MIE). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge MA, Harvard University Press.
- Brandom, R. (1983). 'Asserting.' *Noûs*, 17, 637-650.
- Brennan, G., Eriksson, L., Goodin, R.E, Brennan, S.B.. (2013). *Explaining Norms*. Oxford, Oxford University Press.
- Byrne, A. (1998). 'Interpretivism.' *European Review of Philosophy*, 3, 199-223.
- Båve, Arvid. (2020). 'Semantic dispositionalism without exceptions.' *Philos Stud* 177: 1751–1771.
- Carnap. R. (1931). 'Überwindung der Metaphysik durch logische Analyse der Sprache.' *Erkenntnis*, Vol. 2, 219-241.
- Carroll, L. (1895). 'What the Tortoise said to Achilles.' *Mind, New Series*, Vol. 4, 278-280.
- Cheng, K. (2009): 'Semantic Dispositionalism, Idealization, and *Ceteris Paribus* Clauses.' *Minds & Machines* 19:407–419.
- Clausen, A. (2008). *How can conceptual content be social and normative, and, at the same time, be objective?* Berlin, Boston, De Gruyter.
- Davidson, D. (1994/2005). 'The Social Aspect of Language.' Published in *Truth, Language, and History*. Oxford, Clarendon Press.
- Davidson, D. (1986a). 'A Coherence Theory of Truth and Knowledge.' Published in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. (ed.) LePore, E. New York, Blackwell.

- Davidson, D. (1986b). 'A Nice Derangement of Epitaphs.' Published in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. (ed.) LePore, E. New York, Blackwell.
- Davidson, D. (1968/2001). 'Actions, Reasons, and Causes.' Published in *Essays on Actions and Events: Philosophical Essays Volume 1*. Clarendon Press.
- Dennett, D. (2010). 'The Evolution of "Why?"' Published in *Reading Brandom: On Making It Explicit*. (ed.) Weiss, B. and Wanderer, J. Taylor and Francis Group, ProQuest Ebook Central.
- Dennett, D. (1971). 'Intentional Systems.' *Journal of Philosophy*, 68, 87-106.
- Devitt, M. (2001). 'A Shocking Idea about Meaning.' *Revue internationale de philosophie*, 4, no. 218, 471-494.
- Devitt, M. and Sterelny, K. (1999). *Language and Reality: An Introduction to the Philosophy of Language*. Second Edition. Oxford, Blackwell.
- Devitt, M. (1996). *Coming to Our Senses: a naturalistic program for semantic localism*. Cambridge MA, Cambridge University Press.
- Dreyfuss, H. (1999). 'The Primacy of Phenomenology over Logical Analysis: A Critique of Searle.' *Philosophical Topics* 27, 3– 24.
- Dummett, M. (1993). *Frege: Philosophy of Language, Second Edition*. Cambridge MA, Harvard University Press.
- Dummett, M. (1978). *Truth and Other Enigmas*. London, Duckworth.
- Errázuriz, J.A. (2014). 'The Performative Contradiction as an Argumentative Device: An Analysis of Its Reach and Scope.' *Logique et Analyse*, nouvelle série, vol. 57, no.225, 15-44.
- Fodor, J. (1992). *A Theory of Content and Other Essays*. MIT Press.
- Fossen, T. (2014). 'Politicizing Brandom's Pragmatism: Normativity and the Agonal Character of Social Practice.' *European Journal of Philosophy* 22:3, 371–395.
- Gendler, T. S. (2008). 'Alief and Belief.' *The Journal of Philosophy*, 105, 634– 663.
- Gibbard, A. (2012). *Meaning and Normativity*. Oxford, Oxford University Press.
- Ginsborg, H. 2021. 'Going on as one ought: Kripke and Wittgenstein on the normativity of meaning.' *Mind & Language*. 1–17.
- Ginsborg, H. 2018. 'Normativity and Concepts.' Published in Daniel Star, (ed.), *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press, 989-1014.
- Ginsborg, H. 2012. 'Meaning, Understanding, and Normativity.' Proceedings of the Aristotelian Society Supplementary Volume lxxxvi.
- Ginsborg, H. 2011. 'Primitive Normativity and Skepticism about Rules.' *The Journal of Philosophy*, Vol. 108, No. 5, 227-254.
- Gómez-Torrente, M. (2019). *Roads to Reference: an essay on reference fixing in natural language*. UK, Oxford University Press.

- Glüer, K. and Pagin, P. (1999). 'Rules of Meaning and Practical Reasoning.' *Synthese* 117: 207–227.
- Glüer, K. and Wikforss, Å. (2009). 'Against Content Normativity.' *Mind, New Series*, Vol. 118, No. 469, 31-70.
- Griffin, N. (2018). 'Brandom and the brutes.' *Synthese*, 195, 5521–5547.
- Haaparanta, L. (2019). *Rajan taju: filosofisia esseitä*. Helsinki, Gaudeamus.
- Hattiangadi, A. (2006). 'Is Meaning Normative?' *Mind & Language*, Vol. 21 No. 2, 220–240.
- Hattiangadi, A. (2003). 'Making it Implicit: Brandom on Rule Following.' *Philosophy and Phenomenological Research*, Vol. LXVI, No.2, 419-431.
- Haukioja, J. (2021). 'Thinking about Things and Thinking about Properties.' A presentation in Tampere University philosophy department, December 2021.
- Haukioja, J., Nyquist, M., Jylkkä, J. (2020) 'Reports from Twin Earth: Both deep structure and appearance determine the reference of natural kind terms.' *Mind & Language*, 36, 377–403.
- Hintikka, J. (1962). 'Cogito, Ergo Sum: Inference or Performance?' *The Philosophical Review*, Vol. 71, No. 1, 3-32.
- Johnston, M. (1997). 'Manifest Kinds'. *Journal of Philosophy* 94, 564 – 83.
- Kaplan, D. (1989). 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals.' Published in *Themes from Kaplan*, (ed.) Almog, J., Perry, J., and Wettstein, H. Oxford, Oxford University Press.
- Katz, J. (1990). *The Metaphysics of Meaning*. MIT Press.
- Katz, J. (1981). *Language and Other Abstract Objects*. Oxford, Basil Blackwell.
- Kiesselbach, M. (2020). 'On Being Bound to Linguistic Norms. Reply to Reinikainen and Kaluziński.' *International Journal of Philosophical Studies*, Vol. 28, 4, 527-540.
- Kiesselbach, M. (2012). 'Constructing Commitment: Brandom's Pragmatist Take on Rule-Following.' *Philosophical Investigations* 35 (2), 101–126.
- Knell, S. (2005). 'A deflationist theory of intentionality? Brandom's analysis of *de re* specifying attitude-ascriptions.' *Pragmatics & Cognition* 13:1, 73–90.
- Kripke, S. (1982/WRPL). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge MA, Harvard University Press.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge MA, Harvard University Press.
- Kowalenko, R. (2009). 'How (not) to think about idealisation and *ceteris paribus*-laws'. *Synthese*, 167:183–201.
- Kusch, M. (2006). *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein*. Acumen, ProQuest Ebook Central.
- Lauer, D. (2012). 'Expressivism and the Layer Cake Picture of Discursive Practice.' *Philosophia*, 40, 55-73.

- Laurier, D. (2005). 'Pragmatics, Pittsburgh style.' *Pragmatics & Cognition* 13:1, 141–160.
- Lewis, D. (1983). 'New Work For a Theory of Universals.' *Australasian Journal of Philosophy* 61 (4):343-377.
- Lance, M. & Kukla, R. (2009). *'Yo!' and 'Lo!': The Pragmatic Topography of the Space of Reasons*. Cambridge MA, Harvard University Press.
- Loeffler, R. (2017). *Brandom*. Polity Press, ProQuest Ebook Central.
- Loeffler, R. (2005). 'Normative Phenomenalism: On Robert Brandom's Practice-Based Explanation of Meaning.' *European Journal of Philosophy*, 13:1, 32-69.
- Maddy, P. (1984). 'How the Causal Theorist Follows a Rule.' *Midwest Studies in Philosophy* 9(1): 457–477.
- McDowell, J. 'Motivating inferentialism: Comments on Making It Explicit (Ch. 2).' *Pragmatics & Cognition*, 13:1, 2–40.
- McGinn, C. (1984). *Wittgenstein on Meaning: An Interpretation and Evaluation*. Aristotelian Society Series, vol. 1. Oxford, Blackwell.
- Miller, A. (2010). 'Rule-Following Skepticism.' Published in *The Routledge Companion to Epistemology*, (ed.) Bernecker, S. & Pritchard, D. ProQuest Ebook Central.
- Millikan, R. (1999). 'Historical Kinds and the 'Special Sciences'.' *Philosophical Studies*, 95, 45–65.
- Millikan, R. (1990). 'Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox.' *The Philosophical Review*, Vol. 99, No. 3, 323-353.
- Moyar, D. (2020). 'Intentional Agency and Conceptual Idealism: Brandom on Hegelian reason.' Published in *Reading Brandom: On A Spirit of Trust*. (ed.) Bouché, G. New York, Routledge.
- Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.
- Peacocke, C. (1984). 'Review: Wittgenstein on Rules and Private Language by Saul A. Kripke.' *The Philosophical Review*, Vol. 93, No.2, 263-271.
- Peregrin, J. (2021). 'Normative Attitudes.' Published in *The Social Institution of Discursive Norms : Historical, Naturalistic, and Pragmatic Perspectives*. (ed.) Townsend, L., et al. Taylor & Francis Group, ProQuest Ebook Central.
- Peregrin, J. (2014). *Inferentialism: Why Rules Matter*. UK, Palgrave MacMillan.
- Peregrin, J. (2012). 'Inferentialism and the Normativity of Meaning.' *Philosophia*, 40, 75–97.
- Pesonen, R. *Casual Reasoning: A Social Ecological Look at Human Cognition and Common Sense*. Tampere University Dissertations, No. 96. <http://urn.fi/URN:ISBN:978-952-03-1172-8>
- Pippin, R. (2005). 'Brandom's Hegel.' *European Journal of Philosophy*, 13:3, 381–408.
- Podlaskowski, A. (2012). 'Simple Tasks, Abstractions, and Semantic Dispositionalism.' *dialectica* Vol. 66, No 4, 453–470.

- Price, H. (2013). 'Prospects for global expressivism'. Published in *Expressivism, Pragmatism, Representationalism*, (ed.) by Huw Price. Cambridge MA, Cambridge University Press.
- Price, H. (2004). 'Naturalism Without Representationalism.' Published in *Naturalism in Question*, (ed.) Mario De Caro and David Macarthur. Cambridge MA, Harvard University Press.
- Prien, B. (2010). 'Robert Brandom on Communication, Reference, and Objectivity.' *International Journal of Philosophical Studies*, 18:3, 433-458.
- Prior, A. (1960). 'The Runabout Inference-Ticket.' *Analysis*, 21, 38-39.
- Quine, W. (1951). 'Two Dogmas of Empiricism.' *The Philosophical Review*, Vol. 60, No. 1, 20-43.
- Raatikainen, P. (2013). 'The Semantic Realism/Anti-Realism Dispute and Knowledge of Meanings.' *Baltic International Yearbook of Cognition, Logic and Communication*: Vol. 5., 1-13.
- Reinikainen, J. (2022). 'Meaning in time: on temporal externalism and Kripkenstein's skeptical challenge.' *Synthese* 200, 288.
- Reinikainen, J. (2021a). 'Brandom and the Pragmatist Quest for Semantic Objectivity.' *Acta Philosophica Fennica*, 97, 55-78.
- Reinikainen, J. (2021b). 'Kripkenstein semanttista realismia vastaan.' *Ajatus*, 78 (11), 187-219.
- Reinikainen, J. (2020). 'Meaning Still Not Normative: On Assessment and Guidance.' *International Journal of Philosophical Studies* Vol. 28, Issue 4, 510-526.
- Rosen, G. (1997). 'Who Makes the Rules Around Here?' *Philosophy and Phenomenological Research*, Vol. 57, No.1, 163-171.
- Rosenkranz, S. (2003). 'Pragmatism, Semantics, and the Unknowable.' *Australasian Journal of Philosophy*, 81:3, 340-354.
- Rosenkranz, S. (2001). 'Farewell to Objectivity: A Critique of Brandom.' *The Philosophical Quarterly*, Vol. 51, No. 203, 232-237.
- Rydenfelt, H. (2019). 'Realism without representationalism.' *Synthese*, 198, 2901-2918.
- Rödl, S. (2010). 'Normativity of Mind Versus Philosophy as Explanation: Brandom's Theory of the Mind.' Published in *Reading Brandom: On Making It Explicit*. (ed.) Weiss, B. and Wanderer, J. Taylor and Francis Group, ProQuest Ebook Central.
- Salis, Pietro. (2019). 'Does language have a downtown? Wittgenstein, Brandom, and the game of 'giving and asking for reasons''. *Disputatio. Philosophical Research Bulletin* 8, no. 9, 00–00.
- Searle, J. (1983). *Intentionality*. Cambridge, Cambridge University Press.
- Sellars, W. (1956). 'Empiricism and the Philosophy of Mind.' *Minnesota Studies in the Philosophy of Science* 1:253-329.

- Shapiro, L. (2004). 'Brandom on the Normativity of Meaning.' *Philosophy and Phenomenological Research*, Vol. LXVIII, No. 1, 141-160.
- Soames, S. (1997). 'Skepticism about Meaning, Indeterminacy, Normativity, and the Rule-Following Paradox.' *Canadian Journal of Philosophy* 27: 211-249.
- Tennant, N. (1997). *The Taming of the True*. New York: Clarendon Press.
- Tison, R. (2022). 'The fanciest sort of intentionality: Active inference, mindshaping and linguistic content.' *Philosophical Psychology*, <https://doi.org/10.1080/09515089.2022.2062315>.
- Turner, S. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge and Presuppositions*. UK, Polity Press.
- Turner, S. (2010). *Explaining the Normative*. UK, Polity Press.
- Vogelmann, F. (2020). 'Keep score and punish: Brandom's concept of responsibility.' *Philosophy and Social Criticism*, Vol. 46 (8), 922–941.
- Wanderer, J. (2021). 'An I without a You? An Exercise in Normative Pragmatics.' Published in *The Social Institution of Discursive Norms*, (ed.) Townsend, L., Stovall, P., Schmid, H.B. Routledge.
- Wanderer, J. (2008). *Robert Brandom*. Trowbridge, Cromwell Press.
- Warren, J. (2020). 'Killing Kripkenstein's Monster.' *Noûs* 54:2, 257–289.
- Whiting, D. (2006). 'Between Primitivism and Naturalism: Brandom's Theory of Meaning.' *Acta Analytica*, 21, 3-22.
- Wilson, G. (1998). 'Semantic Realism and Kripke's Wittgenstein.' *Philosophy and Phenomenological Research*, Vol. 58, No. 1, 99-122.
- Wittgenstein, L. (1958/PD): *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford, Basil Blackwell.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. With an Introduction by Bertrand Russell. (ed.) C.K.Ogden. UK, Kegan Paul.
- Wright, C. (1984). 'Kripke's account of the argument against private language.' *Journal of Philosophy* 81 (12):759-778.
- Wright, J. (2012). 'In Defence of Kripkenstein: On Lewis' Proposed Solution to the Sceptical Argument.' *International Journal of Philosophical Studies* Vol. 20(5), 603–621.
- Zalabardo, J. (1997). 'Kripke's Normativity Argument.' *Canadian Journal of Philosophy*, 27:4, 467-488.

