EVA PICARDI ON LANGUAGE, ANALYSIS AND HISTORY

Annalisa Coliva

EVA PICARDI ON LANGUAGE, ANALYSIS AND HISTORY

Edited by Annalisa Coliva, Paolo Leonardi and Sebastiano Moruzzi

Eva Picardi on Language, Analysis and History

Annalisa Coliva · Paolo Leonardi
Sebastiano Moruzzi
Editors

Eva Picardi on Language, Analysis and History

palgrave
macmillan

Editors
Annalisa Coliva
University of California Irvine
Irvine, CA, USA

Paolo Leonardi
Department of Philosophy and Communication Studies
University of Bologna
Bologna, Italy

Sebastiano Moruzzi
Department of Philosophy and Communication Studies
University of Bologna
Bologna, Italy

ISBN 978-3-319-95776-0 ISBN 978-3-319-95777-7 (eBook)
https://doi.org/10.1007/978-3-319-95777-7

Library of Congress Control Number: 2018948815

Cover credit: © A_Pobedimskiy. Getty Images

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction 1
Annalisa Coliva, Paolo Leonardi
and Sebastiano Moruzzi

Part I Themes from Frege

Early Analytic Philosophy's Austrian Dimensions 7
Kevin Mulligan

Truth, Ascriptions of Truth, and Grounds of Truth Ascriptions 31
Wolfgang Künne

The Names of the True 67
Paolo Leonardi

Was Frege a Logicist for Arithmetic? 87
Marco Panza

Logic as Science 113
Robert May

Thin Reference, Metaontological Minimalism and Abstraction Principles: The Prospects for Tolerant Reductionism 161
Andrea Sereni

A Context Principle for the Twenty-First Century 183
Fabrizio Cariani

Slurs and Tone 205
Ernie Lepore and Matthew Stone

Refusing to Endorse: A Must Explanation for Pejoratives 219
Carlo Penco

Fregean Presentationalism 241
Elisabetta Sacchi

Part II Themes from Davidson

Agency Without Rationality 265
Lisa Bortolotti

Reasons and Causes in Psychiatry: Ideas from Donald Davidson's Work 281
Elisabetta Lalumera

The Doxastic Zoo 297
Pascal Engel

Part III Language, Contextualism and Naturalism

Naturalizing Picardi 319
Diego Marconi

Practical Knowledge and Linguistic Competence 337
Annalisa Coliva

A Plague on All Your Houses: Some Reflections on the Variable Behaviour of "Knows" 357
Crispin Wright

Truth Relativism and Evans' Challenge 385
Sebastiano Moruzzi

Knowing the Facts: A Contrastivist Account of the Referential Opacity of Knowledge Attributions* 401
Giorgio Volpe

Index 421

Contributors

Lisa Bortolotti Department of Philosophy, University of Birmingham, Birmingham, UK

Fabrizio Cariani Northwestern University, Evanston, IL, USA

Annalisa Coliva University of California Irvine, Irvine, CA, USA

Pascal Engel EHESS-Ecole des hautes études en sciences sociales, Centre de recherche sur les arts et le langage (CRAL), Paris, France

Wolfgang Künne University of Hamburg, Hamburg, Germany

Elisabetta Lalumera University of Milano-Bicocca, Milan, Italy

Paolo Leonardi University of Bologna, Bologna, Italy

Ernie Lepore Rutgers University, New Brunswick, NJ, USA

Diego Marconi Department of Philosophy and Education Studies, University of Turin, Turin, Italy

Robert May University of California, Davis, CA, USA

Sebastiano Moruzzi Department of Philosophy and Communication Studies, University of Bologna, Bologna, Italy

Kevin Mulligan Università della Svizzera italiana, Lugano, Switzerland

Marco Panza CNRS, IHPST (CNRS and University of Paris 1, Panthéon-Sorbonne), Paris, France; Chapman University, Orange, CA, USA

Carlo Penco University of Genoa, Genoa, Italy

Elisabetta Sacchi Vita-Salute San Raffaele University, Milan, Italy

Andrea Sereni Scuola Universitaria Superiore IUSS Pavia, Pavia, Italy

Matthew Stone Rutgers University, New Brunswick, NJ, USA

Giorgio Volpe University of Bologna, Bologna, Italy

Crispin Wright New York University, New York, NY, USA

Introduction

Annalisa Coliva, Paolo Leonardi and Sebastiano Moruzzi

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_1

Eva Picardi was born in Reggio Calabria, Italy, January 16, 1948. Early after, her family moved to Bologna. At Bologna, shortly after 1970, she graduated in Philosophy. In 1984 she received a D.Phil in Philosophy, presenting a dissertation on *Asserting*, written under Sir Michael Dummett's supervison. Asserting was also the topic of her first book, *Assertibility and Truth / A Study of Fregean Themes* (1981). Eva taught Philosophy of language at the University of Bologna from 1976 until 2016. She died in Bologna, April 23, 2017.

Eva Picardi was one of the promoters of the European Society for Analytic Philosophy; she was on the Editorial Board of the *European Journal of Philosophy* and of the *Journal for the History of Analytic Philosophy*. She was also a member of the *Advisory Board* of the Palgrave Macmillan series on the History of Analytic Philosophy, edited by Michael Beaney, and a member of the advisory board of the group coordinated by Crispin Wright for the translation of Frege's *Grundgesetze der Arithmetik* into English.

Her research was on themes and authors of analytic philosophy of language. She was an internationally renowned expert on Gottlob Frege, whose work she related to, and confronted with that of Giuseppe Peano, Bertrand Russell, Ludwig Wittgenstein, Frank Ramsey, Rudolf Carnap and Donald Davidson. Some of her essays on these topics were collected in the volume *La chimica dei concetti. Linguaggio, logica, psicologia 1879–1927* [*The Chemistry of Concepts—Language, Logic, Psychology 1879–1927*] (1994). After 1990 she began intense research on American neo-pragmatists—W.V.O. Quine, Donald Davidson, Hilary Putnam, Richard Rorty e Robert Brandom—often comparing their work with Michael Dummett's anti-realist program. More recently she got interested in the contextualist debate, tracing its origin back to

A. Coliva (✉)
University of California Irvine, Irvine, CA, USA
e-mail: a.coliva@uci.edu

P. Leonardi
Department of Philosophy and Communication Studies,
University of Bologna, Bologna, Italy
e-mail: paolo.leonardi@unibo.it

S. Moruzzi
Department of Philosophy and Communication Studies,
University of Bologna, Bologna, Italy
e-mail: sebastiano.moruzzi@unibo.it

Frege's distinction between sense and tone, his context principle, later radicalized by Wittgenstein and by Quine's and Davidson's holism—of which Dummett proposed a milder, molecularist variant.

Eva Picardi was also very active in translating analytic philosophers—Frege, Wittgenstein, Davidson, Putnam, Dummett—into Italian. Together with Carlo Penco, she worked on a new Italian edition of Frege's works until her very last days. Her contribution to the development of Italian analytic philosophy is much greater, though. She was the editor of the philosophical section of *Lingua e stile*, from 1992 until 2000, a journal that soon became the most important one for Italian philosophers of language. Besides, she was one of the ten founders of the Italian Society for Analytic Philosophy, and its President from 2000 until 2002. Together with Annalisa Coliva, she organized in 2001 the conference *Wittgenstein together*, which assembled in Bologna the best scholars on Wittgenstein. From the meeting an anthology by the same title originated.

A great teacher, Eva Picardi motivated and directed her students, dedicating a lot of time and energy to them. Many of her students have become researchers and professors in Italy, in other European countries (United Kingdom, Finland, Portugal, Germany), in the United States and New Zealand. Her engagement with teaching made her write one of the most complete introductions to philosophy of language *Linguaggio e analisi filosofica. Elementi di filosofia del linguaggio* (1992) and later a more agile one, *Teorie del significato* (1999), translated into Spanish in 2001.

Her stand on meaning was moderately literalist. She was aware of the role of pragmatic aspects and context in understanding, but she kept to a normative view of meaning. Normativism made her critical of naturalist programs, such as Chomsky's and Fodor's, which she carefully examined. A Fregean, she never became a supporter of direct reference, her admiration for Saul Kripke, Hilary Putnam and Tyler Burge notwithstanding. Philosophy of language was, according to her, first philosophy—that is, the way for investigating the main if not all philosophical topics. At that, she largely endorsed Dummett's view as presented in his *Origins of Analytic Philosophy* (a series of lectures Dummett delivered in Bologna at Eva Picardi's invitation).

Eva Picardi had style—philosophical and personal. She mastered her field and had knowledge beyond it. She was no sceptic, and had firm philosophical convictions and, in discussion, she was precise and insightful. At the same time, she would often not argue the last steps: references and quotations insinuated a different ground and the unfinished argument left the conclusion open. It was lightness and respect, and more.

She was convinced that matters can be seen in more than one way, and that this is what rewards us in a vast knowledge of the literature. A perspicuous picture, which is what we constantly aim at, is one that looks at its object from any of the surrounding points of view, for an indefinite span of time, i.e., an impossible picture. That brings no regret—world and life are richer than any picture of them.

This volume is meant to honour Eva Picardi—her philosophical views and interests, as well as her teaching. It collects eighteen essays, some by former students of hers, some by colleagues with whom she discussed and interacted. The volume has three sections: one of Frege's work—in philosophy of language and logic—, taking into account also its historical dimension; one on Davidson's work; and one on the contextualism-literalism dispute about meaning and on naturalist research programmes such as Chomsky's. The volume reproduces also a picture by Salvatore Nocera, an Italian painter whose work has been shown in an exhibition Eva Picardi prepared during the last couple of years of her life, but did not manage to see. Some of Nocera's paintings adorned her house which she furnished following Adolf Loos' dictates in *Ornament and Crime* (1908).

During her long illness, she neither hid her condition nor turned it into a problem to participate, going on as if there were no deadlines, undertaking various new projects, among which plans for the Summer 2017. Elegant and beautiful, intelligent and cultivated, fearless as she was.[1]

Note

1. If you would like to have a glimpse of Eva Picardi's serious irony, you may watch this video: https://www.youtube.com/watch?v=kiWVa4lIyU4If, you would like to listen to a lecture by her, you may watch this other video: http://www.cattedrarosmini.org/site/view/view.php?cmd=view&id=213&menu1=m2&menu2=m37&menu3=m410&videoid=935.

PART I

Themes from Frege

Early Analytic Philosophy's Austrian Dimensions

Kevin Mulligan

1 Introduction

Allusions are often made to the more or less intimate relations between early analytic philosophy—Frege, Russell, Moore—and Austrian philosophy—Brentano and his students, Meinong, Husserl, Ehrenfels, Twardowski, Marty, Kerry and Stumpf. But my impression is that in spite of the pioneering efforts of Eva Picardi,[1] Roderick Chisholm, Michael Dummett, Peter Simons, Barry Smith and others, these relations are still unfamiliar and ill-understood. The relevant relations are of two kinds. First, the conceptual relations between the philosophies of the Austrians and the philosophies of the founders of analytic philosophy. Second, relations of influence and epistemic relations—the knowledge some of these philosophers had of the philosophies of the others, in particular what they learned from each other. Claims about relations of the second kind often presuppose some grasp of relations of the first kind particularly when they go beyond claims about who read, praised, criticised what.

K. Mulligan (✉)
Università della Svizzera italiana, Lugano, Switzerland
e-mail: Kevin.Mulligan@unige.ch

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_2

In what follows I set out a series of sketches of some aspects of the relations between early analytic and Austrian philosophy.

2 Stout Opens Cambridge's Doors to the Austrians

Stout, a teacher of Russell and Moore, seems to have introduced the philosophies of Brentano and his pupils to Cambridge.[2] Russell and Moore then quickly turn their attention to the ideas mentioned by Stout and to other publications in the Brentanian tradition. And other Cambridge philosophers, Broad, McTaggart and Laird and, for example, Findlay, were to keep this interest in Austrian philosophy alive.[3]

Stout refers in his 1896 *Analytic Psychology* (I, II) to Brentano's *Psychologie* (1874), *Vom Ursprung sittlicher Erkenntnis* (1889) and to his article „Das Genie" (1882). He also refers to the work of three of Brentano's students—Ehrenfels' 1890 article "Über Gestaltqualitäten" (Stout 1896, I 65), Meinong's 1891 article "Zur Psychologie der Komplexionen und Relationen" and to work by Stumpf (Stout 1896, I 56–59, 70–71, 250). Stout discusses in detail Brentano's account of mental modes (Stout 1896, I 38–43), Brentano's arguments in favour of the distinction between presentings, on the one hand, and believing or judging, on the other hand (Stout 1896, I 99–111) and "Brentano's analysis" of the distinction between feeling and conation (Stout 1896, I 116–121). He briefly outlines Brentano's views about intellectual and non-intellectual correctness or rightness, the correctness of judgings and of emotings, in the course of expounding Brentano's distinction between presentings, on the one hand, and judgings, emotings and conatings, on the other hand:

> Brentano points out that mere ideas cannot strictly speaking be right or wrong. They do not possess any virtue or vice, if we may be allowed the expression, by reason of which they can be approved or disapproved of. Within the sphere of desire,—of love and hate,—it is otherwise. Here we find a distinction between the morally good and the morally bad. Similarly, in the case of belief there is a corresponding distinction between truth and error. (Stout 1896, I 111)

As a summary of Brentano's view, this is both incorrect and idiosyncratic but it does present Brentano's main claim, a revival of a view to be found in Plato and Aristotle, that judgings as well as emotings and

desiring—which Brentano calls phenomena of love and hate—are correct or incorrect.[4] Stout correctly notes that "Brentano seems to play with the word *Lieben* (liking), much as Mill does with the word 'pleasure'" (Stout 1896, I 120).

Stout concurs with (what he calls) Brentano's view that affirmation and denial can vary in intensity and goes on to distinguish the degrees of firmness or fixity of assurance and conviction from degrees of intensity of belief. Like some later philosophers, he accepts the former and rejects the latter (Stout 1896, I 110).

Chapter III of Stout's (1896) *Analytic Psychology* is entitled "The Apprehension of Form" and deals with mereology and its epistemology. It begins as follows:

> Every whole involves (i) component parts, and (2) the form of combination in which these parts are united. The nature of the components varies in different cases, and so does their mode of grouping. We have now to consider the following questions : How far is the apprehension of a certain form of combination distinct from and independent of the apprehension of its constituent parts? and, conversely: How far is the apprehension of the components of a certain kind of whole distinct from and independent of the apprehension of its form of synthesis? (Stout 1896, I 65)

These distinctions and questions are at the centre of the work of Brentano and his pupils. And Stout appends an interesting note to the introductory passage just quoted, referring to the already mentioned 1890 article in which Ehrenfels launched Gestalt Psychology:

> Chr. Ehrenfels…has discussed certain aspects of this question …What I designate as form or plan of combination he calls a 'shape-quality'. This use of words sounds strange in German, and it would certainly appear very uncouth in English. I have preferred to say 'form' instead of 'shape'. It is advisable however to point out that my application of the word coincides rather with ordinary usage than with the technical usage of Kant. Form in the text does not stand for the universal and necessary as opposed to the particular and contingent. Forms of combination may be as concrete and particular as the elements combined.

In other words, the forms of combination of particular sounds are as concrete and particular as the sounds themselves, they are all what Stout will later call "abstract particulars" and what Husserl had already called

"Momente". Stout's concrete, particular forms of combination are Husserl's "moments of unity", of which "figural moments" are a species.

Stout distinguishes sharply between mereology and its epistemology:

> It should be noted that we do not propose to investigate the relation of combination to elements combined in the actual constitution of an objective whole; our problem concerns only the relation of apprehension of form to apprehension of matter. Even if an objective whole is nothing more than the sum of its parts taken collectively, it does not follow that our cognisance of this whole is to be identified with our cognisance of all its parts. In the next place it must be understood that we are not here concerned with mental combination. If the apprehension of form is in any sense distinct and independent of the apprehension of matter, it is itself not a form of mental combination, but a material constituent of consciousness, comparable in this respect with the perception of red or blue. In the sequel we shall have to inquire how, in mental process, the apprehension of the form of a whole conditions and is conditioned by that of its constituent parts. This will be in a strict sense an inquiry into mental form of combination. (Stout 1896, I 65–66)

He then asks: "Can the form of combination remain the same or relatively the same, while the constituents vary?" His affirmative answer is a variation on Ehrenfels' account of the transposability of Gestalt qualities, although Ehrenfels is no longer mentioned. But Stout does refer to the discussion of an alternative answer given by Meinong (Stout 1896, I 70). Stout also asks: "Is it possible to apprehend all the components of a whole without apprehending their mode of connection?" and gives an affirmative answer to this question, too, quoting Stumpf's view that

> we may be aware of two notes differing in pitch, and we may be aware that they do so differ, without observing which is higher than the other. (Stout 1896, I 70–71)

Stout agrees with James's rejection of Stumpf's view of psychological analysis, though not with James's arguments and alternative:

> Like us, though on different, and, I think, partly fallacious grounds, he denies that when, by an effort of attention, we transform an indistinct into a distinct presentation, the features distinguished by consciousness in the latter were precontained as real, though undiscerned, components in the

> former. Thus, according to James, it is inaccurate to say, with Stumpf, that by mental analysis we can "clearly perceive that the content of our sensation of oil of pepper mint is partly a sensation of taste and partly one of temperature". He rightly refuses to admit that an original indistinct content of sensation can be legitimately identified with a subsequent distinct one. The explanation which he proposes is that "we perceive the objective fact, known to us as the peppermint taste, to contain those other objective facts known as aromatic or sapid quality, and coldness, respectively". In like manner he would resolve all so-called analysis of presentations into analysis of the objective facts, which are known by means of them. This view is advanced by Professor James with dogmatic emphasis. But he does not support it on positive grounds. He seems to consider that it is adequately established by a refutation of the common doctrine, which I agree with him in rejecting. It apparently does not occur to him that there may be another alternative. It is at this point that we cease to be able to follow him. His doctrine, if pushed to its logical consequences, would involve the impossibility, not merely of the "analysis of presentation" but of all analysis properly so called.[5]

The view of psychological analysis rejected by James and Stout—for different reasons—is the view which will also be rejected by Stumpf's pupils, the Berlin Gestalt Psychologists—Köhler and Koffka—in their criticisms of Stumpf, Ehrenfels and the Graz school (Meinong, Benussi, Witasek) of Gestalt Psychology.[6]

In his 1899 *Manual* Stout sets out a view of quantities and measurement which is based to a large extent on Meinong's (1896) monograph *Über die Bedeutung des Weberschen Gesetzes. Beiträge zur Psychologie des Vergleichens und Messens.* As Stout says: "The treatment of Weber's law in this chapter [ch. 7] follows Meinong, *Ueber die Bedeutung des Weberschen Gesetzes*, etc." (Stout 1899, 209).

Each of the Austrian ideas discussed by Stout was to be discussed by Cambridge philosophers. In 1899 Russell publishes his own detailed review of Meinong's work on quantities and relations and in his 1903 *Principles of Mathematics* (Chapters XIX–XX, XXII) draws on Meinong's philosophy of quantities and relations. Of Meinong's monograph he says that he has "learnt so much" from it and that he "largely agrees with" it (Russell 1979, 168). In 1903, Moore discusses at length Brentano's account of correct or right emoting and preferring and his account of goodness. Indeed Russell was to go on to review many works by Meinong and his pupils (Russell 1904, 1905a, b, 1906, 1907). And the

distinction drawn by Ehrenfels and Stout between the parts of a whole and their forms of combination was to become of great importance in Cambridge philosophy, in particular, as the distinction between the parts of a whole and the relations between the parts, in the writings of Wittgenstein. The relations between the parts of a whole, Husserl and then Wittgenstein argue, are not parts of the whole. The epistemology of wholes, their parts and forms of combination, the nature of perception and knowledge of these, was also to become a preoccupation of Cambridge philosophers. But in early Austrian philosophy, as we have seen, the analysis of wholes meant in the first place the analysis of perceived wholes and the analysis of mental acts and states. Examples of these kinds quickly become less important in Cambridge philosophy than the analysis of the wholes called propositions. In Austrian philosophy, it is first of all in the writings of Husserl that the analysis of propositions and meanings takes centre stage. And both Husserl and Meinong propose philosophies of the wholes they call states of affairs, *Sachverhalte* (Husserl), and objectives (Meinong) and which Russell sometimes calls propositions and sometimes objectives.

This brief and selective account of the way in which certain *questions and answers thereto* seem to have arrived in Cambridge would be incomplete without an account of the *way* in which Cambridge philosophers reacted to the way in which Austrian philosophers did philosophy and the relation between this model and what was to become the most distinctive trait of early, Cambridge analytic philosophy—the obsession with and search for clarity.

3 Clarity and Cognitive Virtue—How to Do Austrian, Analytic Philosophy

There is a definite and distinctive continuity between the tone and type of Stout's praise for the way the Austrians do philosophy, on the one hand, and the praise Russell and Moore were to lavish on Meinong and Brentano in particular.

Thus in 1896 Stout says that the question of how to distinguish between presentings and judgings is "treated with admirable care and acuteness" by Brentano (Stout 1896, I 99). Ehrenfels, he says, "has discussed" certain aspects of the relation between the apprehension of the form of a whole and apprehension of its parts "with great fullness and

precision" (Stout 1896, I 65). Of an argument by Stumpf in favour of the thesis "that sensation can actually exist without cognitive function" (Stout 1899, 120), he says that its "merit" "lies in the exact and cogent form into which it is thrown" (Stout 1899, 121).

When Russell and then Moore turn their attention to the writings of Brentano and Meinong described by Stout, they praise what Stout had praised but more insistently. In Russell's 1899 review of Meinong's monograph on quantity and measurement, which Stout summarises and discusses in the same year, he says:

> The present work consists essentially of a single thesis proved by a single argument. The thesis is at once simple and ingenious, the argument at once lucid and subtle. The author avoids almost all the mistakes and confusions which beset writers on psychical measurement, and makes several important distinctions which are rarely, if ever, to be met with elsewhere. (Russell 1899)

In 1903 Russell says of the belief that all order depends on distance that, "though entertained by so excellent a writer as Meinong", it is false (Russell 1903, 419). In 1907 Russell writes of a monograph published in the same year by Meinong that "the style is remarkably clear", that Meinong's "contentions are in all cases clear" and that

> the polemical arguments appear to the present reviewer to be generally cogent, except (needless to add) when they are directed against himself. (Russell 1907)

Moore's discussions of the 1902 English translation of Brentano's 1889 *Vom Ursprung sittlicher Erkenntnis* and in particular of just the ideas summarised by Stout in 1896, are published in 1903: a review and a page in Moore's Preface to *Principia Ethica*. The latter tells the reader:

> When this book had already been completed, I found, in Brentano's 'Origin of the Knowledge of Right and Wrong', opinions far more closely resembling my own than those of any other ethical writer with whom I am acquainted. (Moore 1966, x–xi)

The review is a tribute from one master-clarifier to another. Moore refers to "Brentano's extraordinary clearness with regard to the precise

relevance of all he says" and says that "Brentano is both clearer and more profound" than Sidgwick. Brentano's "is a far better discussion of the most fundamental principles of ethics than any others with which I am acquainted".

Moore's longest *Auseinandersetzung* with the Brentanian tradition is his 1910 review of a primer of Husserl's philosophy of mind and language by the German philosopher and psychologist August Messer (1908), although he occasionally refers later to both Meinong and Stumpf. Moore says of Messer's book that it is "extraordinarily good" and "written beautifully simply and clearly"; the author is "wonderfully successful in making plain, by means of examples, exactly what it is that he is talking about" (Moore 1910).

The praise which Russell and Moore bestow on Meinong was also given by younger Cambridge philosophers. Thus Broad concludes his 1913 review of the second edition of Meinong's *Über Annahmen* with words which echo Russell's review of the first edition: "The book as a whole can safely be described as a model of acute and profound investigation into the hardest and most fundamental questions of philosophy".

But perhaps the most striking example of the effect of exposure to Austrian methodology in Cambridge is to be gleaned from a comparison of Russell's 1911 account of the true method, in philosophy and science, analytic realism and logical atomism, with his 1904 account of Meinong's way of doing philosophy. In 1911 Russell writes:

> There have been far too many heroic solutions in philosophy; detailed work has too often been neglected; there has been too little patience. As was once the case in physics, a hypothesis is invented, and on top of this hypothesis a bizarre world is constructed, there is no effort to compare this world with the real world. The true method, in philosophy as in science, will be inductive, meticulous, and will not believe that it is the duty of every philosopher to solve every problem by himself. This is the method that inspires analytic realism and it is the only method, if I am not mistaken, by which philosophy will succeed in obtaining results which are as solid as those of science. (Russell 1911, 61)

In 1904, in a review of Meinong's monographs *On Assumptions* and *On Higher-Order Objects*, he had written:

> Although empiricism as a philosophy does not appear to be tenable, there is an empirical method of investigating, which should be applied in every

> subject-matter. This is possessed in very perfect form by the works we are considering. A frank recognition of the data, as inspection reveals them, precedes all theorising; when a theory is propounded, the greatest skill is shown in the selection of facts favourable or unfavourable, and in eliciting all consequences of the facts adduced. There is thus a rare combination of acute inference with capacity for observation. The method of philosophy is not fundamentally unlike that of other sciences: the differences seem to be only in degree…Whatever may ultimately prove to be the value of Meinong's particular contentions, the value of his method is undoubtedly very great; and on this account if on no other, he deserves careful study. (Russell 1973, 22–23)

Wittingly or unwittingly, Russell here echoes the thesis which Meinong's teacher, Brentano, had defended during his *Habilitation* in 1866: "the true method of philosophy is none other than that of the natural sciences."

The different features of the Austrian way of doing philosophy which Stout, Russell and Moore single out for praise are simply some of the different components or aspects of cognitive virtue, with a definite preference for the striving for clarity which Bolzano and then Brentano had introduced into Austrian philosophy.

The philosophical questions mentioned so far, as they were posed and answered by Brentano and his heirs and then taken up by Cambridge philosophers have come in for some discussion.[7] In what follows, I consider two less well-known Austro-Cambridge ideas.

4 The Present Emperor of France vs. the Present King of France

Russell's 1905 account of definite descriptions appeared four years after Husserl's inconclusive reflexions on the same subject. But the relations between the views of Husserl and Russell on reference and denotation are of some interest and not just to the historian of early analytic and Austrian philosophy. Although Husserl's early views about definite descriptions are not worked out very clearly, he has a sophisticated and plausible account of proper names and what he calls "occasional", that is to say, demonstrative and indexical expressions, singular and plural. Unlike Russell and Carnap he knew that these two categories are very different. Roughly, his view is that (singular) proper names have a sense

or meaning which, in the most basic cases, is simple and so non-descriptive and that the object of such a name, if it has one, is not its meaning or sense. The reference of a proper name is fixed by non-conceptual perception or by something which plays the same role. But someone who does not understand the proper meaning of a proper name, for example "Madrid", may still use it correctly provided he has learnt that, as Husserl puts it, "the capital of Spain is called (has the proper name) Madrid" (Husserl 1984b, VI §5, 556). In such a case, we may say, the reference of the proper name has been fixed by a definite description. Proper names, Husserl says, may refer or name directly or indirectly. Similarly, demonstratives also have a simple sense but a simple sense which, unlike the sense of a proper name, is incomplete and can only be completed by a perceptual content or something which plays the same role as this. Definite descriptions, which Husserl often calls "attributive names", have a meaning or sense which is not simple and they name or refer "indirectly".[8] Husserl would almost certainly not have called attributive names descriptions because he thought, very plausibly, that to describe was to predicate on the basis of perception or of something which plays a similar role. He distinguishes between attributive names for real or temporal, what he also calls "individual" objects and for ideal objects. His examples of the former include

> the victor of Jena
> the loser of Waterloo
> the present Emperor of Germany
> the greatest German statesman
> the lamp. (Husserl 1984a, I §12)

and in a manuscript from 1899 Husserl gives the example of "the Emperor of France" (Husserl 2009, 139). His examples of the latter include

> the equilateral triangle
> the equiangular triangle
> the second even number in the number series. (Husserl 1984a, I §12, §33)

One important difference between the two types of attributive names, he argues, is that definite descriptions for real, temporal entities are *also* occasional (demonstrative, indexical) expressions:

> The essentially occasional character naturally spreads to all expressions which include these and similar presentations as parts…[A]ll combinations involving *the definite article*, in which the latter relates to something individual [temporal] which is only related to by means of class-concepts or property-concepts belong here. When we Germans speak of *the* Kaiser we of course mean the present German Kaiser. When we ask for *the* lamp in the evening, each man means his own. (Husserl 1984a, I §26)

Since "present" and "own" are occasional expressions, Husserl is right to say that, at least in his two examples, definite descriptions for real objects are also occasional.

The definite article common to both types of definite descriptions is, Husserl says, a "formal word", as are "one", "some", "and", "or" and "which" (Husserl 1984b, VI §40). But, as far as I can see, he nowhere gives an account of the relation between definite descriptions, identity and quantification. Two aspects of his views are of interest, in the light of the subsequent discussion of Russell's views. One has to do with Husserl's account of what a speaker (judger) does when he combines an attributive name with a predicate in the context of an assertion (judgment). His interest in this topic is already evident from his claims in the last passage quoted about what a speaker means when he employs a certain sort of expression. This topic is central in the reactions of Strawson and Searle to Russell's account of definite descriptions. The second aspect of Husserl's views concerns not pragmatics or psychology but the relation between propositions containing the meaning of an attributive name and existential propositions.

What is it to mean someone or something with the help of an attributive name? Husserl says in 1899:

> If I say 'the Emperor', I state nothing, i.e. I predicate nothing (in the expression taken by itself). But it 'lies therein' that a real person is involved….Predication is what we consider to be the basic act in logic…If someone says 'The Emperor of France', we object: 'You believe that there is an Emperor of France'. (Husserl 2009, 139)

Unlike Frege and many later friends of the distinction between modes or force and content, Husserl thinks that the distinction applies not just to propositional contents but also to non-propositional parts of such contents. In particular, he thinks that in judging that Sam is sad or in

judging that the lamp is beautiful the act of "meaning" Sam or a lamp involves a non-propositional mode of *positing*. Sam and the lamp are meant in a positing but non-judgmental way, just as the act of meaning that Sam is sad is qualified in a positing, judgmental way. Similarly, in supposing that Sherlock Holmes prefers whiskey to cocaine the act of meaning that Sherlock Holmes prefers whiskey to cocaine is coloured by the non-positing mode of supposing but the act of meaning Sherlock is itself coloured by a non-positing mode.

Moore was familiar with this view, in Messer's presentation of it. And he is again it:

> Dr. Messer supposes that 'propositional' Acts are not the only kind of Acts which can differ from one another in this way [the way in which judging and supposing differ]: he supposes that 'nominal' Acts also can be 'positing' …In this, however, I cannot help thinking he is wrong. So far as I can see, it is not possible to believe anything but a proposition. Dr. Messer only gives as an instance of the cases where, according to him, a ' positing' nominal Act occurs, what happens when we believe such a proposition as "The Emperor Charles conquered the Saxons". When we "posit" this proposition, we also, he thinks, "posit" its subject, the Emperor Charles. But surely there is a confusion here. When we believe such a proposition as this, it is, I think, generally true that we believe also in the existence of the subject; and similarly in propositions about what Dr. Messer calls 'ideal' objects, we generally believe in the 'being' of their subjects, though not in their 'existence'. But surely these beliefs in the existence or the being of a subject are 'propositional' Acts; and I can see no reason 'to think that any further 'positing' Act is involved - a 'positing' Act, for instance, of which the Emperor Charles himself, and not merely his existence, is the object. I am inclined to think, therefore, that Dr. Messer only thinks that nominal 'Acts' can be 'positing,' because he mistakes for a nominal Act, in these instances, what is, in reality, a 'propositional' Act. (Moore 1910)

Messer and Husserl might respond as follows. It is indeed possible to believe things other than propositions, as when Russell believes Wittgenstein. Similarly, if belief in is a type of belief, there is belief in things other than propositions—there is a (non-axiological) belief in atoms as well as the (axiological) belief in capitalism. And belief that *p* is not belief in the proposition that *p*. Indeed, the category of belief in is one way of glossing the idea of a non-propositional, positing mode.[9] What Moore calls beliefs in the existence of a subject are either

dispositional beliefs or so called occurrent beliefs. The specification of a dispositional belief does not belong to an account of the make-up of a mental episode. If the belief in the existence of a subject is an occurrent belief, then one who judges that the Emperor Charles conquered the Saxons judges two things rather than one thing. Whether or not one who sincerely asserts that the Emperor Charles conquered the Saxons presupposes and believes that he exists, the category of positing, non-propositional acts is in any case required to deal with sincere assertion of negative existentials.[10] And also with a case Moore does not discuss (in his review): what Husserl calls simple seeing, as opposed to seeing that, seeing Bismarck as opposed to seeing that he is over there.

Husserl's main claim in 1901 about the relation between non-existential propositions containing positing names—proper names or attributive names—and existential propositions goes as follows:

> That a proposition with some positing names holds and that the existential judgments (*Seinsurteile*) corresponding to these names do not hold is an apriori incompatibility. It belongs to the group of 'analytic ideal laws', which are grounded in 'the mere form of thinking', in the categories… which belong to the possible forms of 'genuine' thinking. (Husserl 1984a, V §35)

In the discussion which precedes this formulation, Husserl indicates that the laws he has in mind belong to his account of semantic modification. According to this account, some meanings are modifications of other meanings and expressions, too, stand in an analogous relation to each other. In particular, a proposition containing a nominalisation of another proposition is a modification of it. Thus Husserl says of examples like

1. Snow is white
2. The proposition that snow is white is true

that (1) and (2) are not merely equivalent but that (2) is a modification of (1) because its nominal component is a nominalisation of (1). He also says that the equivalence between (1) and (2) and the relation of modification between them are grounded in the nature of meanings. His formulations sometimes suggest a claim like that endorsed by Bolzano in his account of grounding (*Abfolge*):

3. If (2), then (2) because (1),

although, as far as I can see, Husserl never clearly says this.

Husserl applies his account of modification to the relation between attributive names and propositions containing attibutive names, on the one hand, and propositions which do not contain such names, on the other hand:

> Without any doubt, many names, including all attributive names, have 'arisen' directly or indirectly out of judgments, and accordingly refer back to judgments. But such talk of 'arising' and 'referring back' implies that names and judgments are different. The difference is so sharp, that it should not be played down for the sake of theoretical prejudice or hoped-for simplifications in the theory of presenting and judging. The prior judging is not as yet the nominal meaning that grows out of it. What in the name is given as a residue of judgment is not a judgment but a modification sharply differing from it. The carrying out of the modified act no longer contains the unmodified one. (Husserl 1984b, V §35)

Husserl's examples of the relation of modification between attributive names (or propositions containing such names) and propositions which do not contain such names fall into two groups. His first type of example includes the relation between the meaning of "The town Halle is on the Saale" and the meaning of "the town Halle on the Saale" (*die Saalestadt Halle*); the latter, he says, is a modification of the former. Similarly, the meaning of "the transcendent number π" is a modification of the meaning of "π is a transcendent number". If we take seriously the analogy with (3) above, we might say that

4. If the transcendent number π is interesting, then the transcendent number π is interesting *in part because* π is a transcendent number.

Husserl's second type of example is the important one for present purposes. It concerns the relation between propositions containing attributive names and the corresponding existential propositions. Husserl's formulations are elliptic and sparse. Unfortunately his clearest formulation of the point he wants to make is in the language of what one can reasonably say:

> one cannot reasonably start with the words "this *S*" without 'potentially' conceding that there are *S*'s. (Husserl 1984b, §35)

The continuation of this passage is the claim, already quoted, that there is an apriori incompatibility between the holding of a proposition containing positing names and the non-holding of an existential judgment corresponding to these names. It is striking that Husserl here at no point wonders, as Russell will, what relation between existential propositions, uniqueness and identity is required to understand propositions containing attributive names.[11] But it would perhaps be in the spirit of Husserl's approach to say of

5. The F is G
6. There is exactly one F and it is G

that

7. If (5), then (5) because (6).

In a passage written before he adopts the point of view set out in his *Investigations*, Husserl formulates a version of the sort of view Russell will endorse:

> The definite article indicates…that the extension of the concept contains only one object, that the concept is therefore a singular concept. It thus expresses an independent judgment to this effect and to which, if the judgment is correct, there corresponds objectively a truth. (Husserl 2001, 78)

In a letter to Husserl (19. 4. 1920), Russell says politely that he has "of course, followed" Husserl's "work with interest and sympathy for many years" and that he had the second edition of Husserl's *Logische Untersuchungen* with him in prison. Whether or not Russell had in fact read any part of Husserl's *oeuvre*, there is little doubt that many of the questions which preoccupied early Russell had also preoccupied and continued to preoccupy Husserl. Some of these questions were relative newcomers to philosophy, such as the relation between identity and substitutability *salva veritate* and (what has been called) Hume's principle, as set out by Frege, whose account is criticized by Husserl in his 1891 *Philosophie der Arithmetik* long before Russell had heard of Frege.

Husserl's most telling criticisms all appeal to the notion of grounding or explanatory priority (*because*), a notion neither Frege nor Russell took very seriously.[12] One old question about which both Husserl and Russell wrote is the right way to defend universals. And, as has been pointed out, their arguments (in the second of Husserl's *Investigations* and in Russell's *The Problems of Philosophy*) are not dissimilar. The questions about the nature of definite descriptions and their meaning or sense and the differences between proper names, occasional expressions and descriptions were introduced into philosophy in the Brentanian tradition by Husserl, under the influence of Bolzano. But these semantic questions are the counterpart of a question which preoccupied Brentano and many of his heirs (and indeed Bolzano): how is the difference between proper, direct, authentic, for example, perceptual or intuitive presentings of an object and more or less indirect or improper presentings of the same object to be understood? This question, which belongs to both epistemology and the philosophy of mind, is central in the early writings of Husserl and Russell, and, too, of Meinong. Broad, in his already mentioned review of Meinong's *Über Annahmen* notes that its author

> distinguishes two kinds of intending. You may intend an object not merely by supposing or entertaining the objective that it exists or subsists, but by doing the like with objectives that assert qualities of it. He calls the former *Seinsmeinen* and the latter *Soseinsmeinen*. We may call them direct and indirect intending respectively. So far as I can see indirect intending corresponds closely to what Mr. Russell calls knowledge of description. (Broad 1913)

But the pioneering accounts of the relation between knowledge by acquaintance and merely descriptive knowledge seem to be those given by Husserl in the fifth of his *Investigations* in 1901 and then by Russell in 1905, 1910 and in particular in chapter five of his 1912 *The Problems of Philosophy*. Common to both accounts is the idea that there are degrees of acquaintance, that as Russell puts it in 1912, "there are various stages in the removal from acquaintance with particulars", that, as both Husserl and Russell say, there is not merely an absolute distinction between direct and indirect epistemic access to objects but a relative distinction between "more or less direct" modes of contact. Husserl refers to stages and levels of epistemic access, to a "phenomenology of levels of knowledge (*Erkenntnisstufen*)", Russell to stages and hierarchies thereof. They are ordered by what Husserl calls internal relations of more or less (*Steigerungsreihen*) (Husserl 1984a, V §§16–25).

5 Meinong's 1906 Proto-picture Theory

Can, should representation be understood in terms of a type of similarity, identity or correspondence between relations, of structural similarity or isomorphism? In Cambridge philosophy, in particular in the writings of Russell and Wittgenstein, affirmative answers to such questions are given. Wittgenstein famously combines the notion of structural similarity with the notion of a picture (*Tractatus* 2.131, 2.15). Fraser MacBride (2018, Chapter 9) draws attention to Russell's 1907 anticipation of picture cum isomorphism theories of representation and refers to Russell's 1906–1907 "Proto Picture Theory".

Meinong's Proto Picture theory seems to have (just) anticipated Russell's version and was published in a monograph reviewed by Russell, although Russell does not clearly refer to this element of Meinong's theory in his review. Neither Meinong in 1906 nor Russell in 1906–1907 actually refer to representations as pictures. But both appeal to similarity or identity between relations in order to understand one type of representation.

Meinong's *Über die Erfahrungsgrundlagen unseres Wissens* was published in 1906 and reviewed by Russell in *Mind* in the same year. One claim made by Meinong is not explicitly mentioned by Russell in his—as usual—positive review: that there are relations which hold both of the noumenal and of the phenomenal properties of things. External perception, which Meinong takes to be a type of judgment, requires identity of relations or relations of the same kind. And at the end of his 1906–1907 paper "On the Nature of Truth", Russell "suggest(s)" the "possibility" of a theory of belief, judgment and perceptual judgment, which appeals to a "correspondence" between relations.[13]

Grossmann summarises Meinong's claims as follows:

> According to Meinong, we perceive the noumenal world since and as far as it shares the same structure with the phenomenal world. That there is shared structure is the unexamined and unquestioned postulate of Meinong's theory of perception.[14]

The "projectible (*übertragbare*, transferable) relations" Meinong appeals to are all formal relations. Although he calls these relations "relations of comparison", he takes relations to be as little mental as Russell: "the difference" between two objects "is not created by comparison but merely grasped with the help of comparison. Were we unable to compare,

we would of course not be able to apprehend the difference; it would nevertheless obtain" (Meinong 1973, 467).

Meinong introduces his account of a shared structure between the phenomenal and the noumenal in the context of a distinction between two types of *Evidenz*. He distinguishes between the *Evidenz* for certainty and conjectural evidence, evidence for surmise (*Vermutungsevidenz*). Memory and external perception provide only *Vermutungsevidenz*. External perception

> provides very good *Vermutungsevidenz* for the existence of things, very bad *Vermutungsevidenz* for the existence of appearing properties (*erscheinenden Eigenschaften*). But the latter are called phenomenal properties because there is also good evidence that things have noumenal properties of which the same relations of comparison hold as hold of the phenomenal properties. (Meinong 1973, 479)

The higher-order objects Meinong calls relations of comparison—numerical identity and difference, equality, similarity and dissimilarity—and number are the two objects which are projectible, transferable, from the phenomenal to the noumenal. They have this capacity because of their "unrestricted applicability", what one might call their formal, categorial or topic-neutral nature. Are non-formal or material higher-order objects, such as melodies, spatial Gestalt and movement projectible? Not directly. But indirectly. This is because material, higher-order objects involve—although they are not built out of—relations of comparison such as difference (Meinong 1973, 470–471).

Meinong, then, appeals to sameneness of relations in an account of perception, which he takes to be a type of judgment. Russell's focus in his 1906–1907 paper is belief and judgment, perceptual and non-perceptual, and Russell, like Husserl in 1900–1901, and unlike Meinong, sharply distinguishes between perception and perceptual judgments. The view of belief Russell explores but does not endorse is this:

> a belief, if this view is adopted, will not consist of one idea with a complex object, but will consist of several related ideas. That is, if we believe (say) that A is B, we shall have the ideas of A and of B, and these ideas will be related in a certain manner; but we shall not have a single complex idea which can be described as the idea of "A is B." A belief will then differ from an idea or presentation by the fact that it will consist of several interrelated ideas. Certain ideas standing in certain relations will be called the

> belief that so-and-so. In the event of the objects of the ideas standing in the corresponding relation, we shall say that the belief is true, or that it is belief in a fact. In the event of the objects not standing in the corresponding relation, there will be no objective complex corresponding to the belief, and the belief is belief in nothing,.... (Russell 1906–1907, 46)

Russell then points out that this view makes it possible to distinguish between perception and judgment based on perception:

> The view that a belief is a complex of ideas, not a single idea, has the merit of distinguishing between the perception of a fact and the judgment which affirms the same fact. We may look at the sky and perceive the sun shining; we may then proceed to judge that "the sun is shining." The same fact, in this case, is first perceived and then judged; the question is: How can the perception and the judgment differ? We may reply that, in the perception, the actual fact or objective complex is before the mind, i.e., there is a single state of mind which has the said objective complex for its object, while in the belief, there is merely a complex of presentations of constituents of the objective complex, these presentations being related in a manner corresponding to that in which the constituents of the objective complex are related. This distinction between perception and judgment is the same as the distinction between intuition and discursive knowledge. The above theory has the merit of explaining, the puzzling fact that perceptions, though they are not judgments, may nevertheless give grounds for judgments. (Russell 1906–1907, 47)

Thus according to Meinong and Russell true perceptual judgments require sameness of relations,[15] although for Meinong but not for Russell, perceptual judgment is perception.

Meinong's intellectual closeness to early Cambridge philosophy was quite evident to one philosopher who knew well the Brentanian tradition. William James refers to "the unspeakable Meinong and his English pals".[16] With equal (in)justice, he might have referred to "the unspeakable Meinong and Husserl and their English pals".[17]

Notes

1. Cf. Picardi (1991, 1994).
2. Cf. the illuminating study van der Schaar (2013); on the differences between Brentano and Stout on judgment, cf. van der Schaar (2013, 66–71). On Brentano, Stout and Moore, cf. Schaar (2017). Cf. also Passmore (1966).

3. Laird (1936) is a Cambridge view of Brentano's influence which resembles the view Brentano's pupil, Oskar Kraus, had of the Brentano effect.
4. Cf. Mulligan (2017a, b).
5. Stout (1896, I 56); James *Psychology*, Vol. I, p. 523, note. That James in fact denies the possibility of psychological analysis is one of the main criticisms in the early (1892), long review of James's *Psychology* by yet another student of Brentano, the Swiss philosopher, Anton Marty (1916). Dennis Seron has suggested to me that the paradox of analysis of psychological wholes has some features in common with the paradox of the analysis of propositions.
6. See the judicious remarks by Mace about Stout's anticipations of later Gestalt Psychology, by which he means especially the psychology of Köhler and Koffka, in his Introduction to the fourth edition of Stout's *Manual*, revised by Mace (1929, ix–x), cf. Stout (1929, 45, 219).
7. On Meinong and Russell on quantities, cf. Guigon (2005); on Moore's review of Messer, cf. Künne (1990); on mereology and its epistemology in Austrian and Cambridge philosophy, cf. van der Schaar (2013), Mulligan (2012, 87–100).
8. Cf. Mulligan (1997).
9. For a discussion of recent attempts to defend this sort of view, cf. Mulligan (2013).
10. Husserl's example in 1901 is "a negative existential judgment such as that 'a triangle with two right angles – does not exist'". In the second edition, the claim is weakened. Husserl says merely that every deliberation about existence requires the category of non-positing names (Husserl 1984b, V §34).
11. In a passage written before he adopts the point of view set out in his *Investigations*, Husserl formulates a version of the sort of view Russell will endorse: "The definite article indicates…that the extension of the concept contains only one object, that the concept is therefore a singular concept. It thus expresses an independent judgment to this effect and to which, if the judgment is correct, there corresponds objectively a truth" (Husserl 2001, 78).
12. Cf. Mulligan (2018b). Three of the earliest, philosophical critiques of Frege were penned by students of Brentano: Marty, Kerry (cf. Picardi 1991, 1994) and Husserl (cf. Mulligan 2018b).
13. Although Russell does not here refer to Meinong's claim about identity of relations, he refers to a paper by Meinong's student, Ameseder, on the perception of complexes.
14. Grossman (1974, 149), on Meinong's theory, cf. Grossman (1974, 143–149).

15. On appeals to structural similarity in the philosophy of science before Meinong and Scheler's criticisms of them, cf. Mulligan (2018a). In 1908 Anton Marty put forward an account of ideal similarity between true judgings and obtaining states of affairs in his *Untersuchungen*. For the suggestion that Marty's ideal similarity is isomorphism, cf. Mulligan (1990, 18–19); on Marty's view, cf. Cesalli and Taieb (2012), Cesalli (2017). It is perhaps no accident that what is perhaps the first brief sketch of the history of uses of isomorphism to understand representation is given by a member of Marty's Prague group, Bergmann (1936).
16. Perry (1935, II 485).
17. Thanks to Paolo Natale, Peter Simons and Maria van der Schaar for their help.

References

Bergmann, H. 1936. Zur Geschichte und Kritik der isomorphen Abbildung. In *Actes du Congrès international de philosophie scientifique*, vol. VII, 65–72. Paris: Hermann.

Brentano, F. 1902. *The Origin of the Knowledge of Right and Wrong*. Westminster: Archibald Constable.

Broad, C.D. 1913. Critical Notice of A. Meinong, *Über Annahmen* (Leipzig, 1910). *Mind* 22 (January): 90–102.

Cesalli, L. 2017. Mental Similarity: Marty and the Pre-Brentanian Tradition. In *Mind and Language—On the Philosophy of Anton Marty*, ed. G. Fréchette and H. Taieb. Berlin: de Gruyter.

Cesalli, L., and H. Taieb. 2012. The Road to *ideelle Verähnlichung*. Anton Marty's Conception of Intentionality in the Light of Its Brentanian Background. *Quaestio* 12: 25–87.

Grossman, R. 1974. *Meinong*. London: Routledge & Kegan Paul.

Guigon, G. 2005. Meinong on Magnitudes and Measurement. *Meinong Studies* 1: 255–296.

Husserl, E. 1984a. *Logische Untersuchungen*, *Vol. II, 1, Husserliana XIX/1*. The Hague: Nijhoff.

Husserl, E. 1984b. *Logische Untersuchungen*, *Vol. II, 2*, *Husserliana XIX/2*. The Hague: Nijhoff.

Husserl, E. 2001. *Logik. Vorlesung 1896*, *Materialienbände, I*. Dordrecht: Springer.

Husserl, E. 2009. *Untersuchungen zur Urteilstheorie. Texte aus dem Nachlass (1893–1918)*, *Husserliana XL*, ed. R. Rollinger. Dordrecht: Springer.

Künne, W. 1990. The Nature of Acts—Moore on Husserl. In *The Analytic Tradition*, ed. D. Bell and N. Cooper. Oxford: Oxford University Press.

Laird, J. 1936. *Recent Philosophy*. London: Butterworth.

MacBride, F. 2018. *On the Genealogy of Universals*. Oxford: Oxford University Press.
Mace, C.A. 1929. *An Introduction to the Fourth Edition* (Stout 1929), vi–xii.
Marty, A. 1916 (1892). Anzeige von William James' Werk: The Principles of Psychology. In *Gesammelte Schriften* I, vol. 1, ed. Josef Eisenmeier, Alfred Kastil, and Oskar Kraus, 105–156. Halle: Niemeyer.
Meinong, A. 1896. *Über die Bedeutung des Weberschen Gesetzes. Beiträge zur Psychologie des Vergleichens und Messens*. Hamburg and Leipzig: Leopold Voss.
Meinong, A. 1906. *Über die Erfahrungsgrundlagen unseres Wissens, Abhandlungen zur Didaktik und Philosophie der Naturwissenschaft*, Band I, Heft 6. Berlin.
Meinong, A. 1907. *Uber die Stellung der Gegenstandstheorie im System der Wissenschaften*. Leipzig: Voigtländer.
Meinong, A. 1973. Über die Erfahrungsgrundlagen unseres Wissens. In *Gesamtausgabe*, ed. R. Chisholm, 367–481. Graz: Akademische Druck- u. Verlagsanstalt.
Messer, A. 1908. *Empfindung und Denken*. Leipzig: Quelle & Meyer.
Moore, G.E. 1903 (Review of Brentano 1902). *International Journal of Ethics*, 14 (October): 115–123 (reprinted in ed. L. McAlister. 1976. *The Philosophy of Brentano*, 176–181. London: Duckworth).
Moore, G.E. 1910 (Review of Messer 1908). *Mind*, New Series, 19 (75, July): 395–409.
Moore, G.E. 1966 (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
Mulligan, K. 1990. Marty's Philosophical Grammar. In *Mind, Meaning and Metaphysics: The Philosophy and Theory of Language of Anton Marty*, ed. K. Mulligan, 11–28. Dordrecht: Kluwer.
Mulligan, K. 1997. How Perception Fixes Reference. In *Sprache und Denken* [Language and Thought], ed. Alex Burri, 122–138. Berlin and New York: de Gruyter.
Mulligan, K. 2012. *Wittgenstein et la philosophie austro-allemande*. Paris: Vrin.
Mulligan, K. 2013. Acceptance, Acknowledgment, Affirmation, Agreement, Assertion, Belief, Certainty, Conviction, Denial, Judgment, Refusal & Rejection. In *Judgement and Truth in Early Analytic Philosophy and Phenomenology*, ed. M. Textor, 97–137. London: Palgrave Macmillan.
Mulligan, K. 2017a. Brentano's Knowledge, Austrian Verificationisms, and Epistemic Accounts of Truth and Value. *Monist* 100 (1): 88–105, special number on Brentano, Guest editor, U. Kriegel.
Mulligan, K. 2017b. Incorrect Emotions in Ancient, Austrian & Contemporary Philosophy. *Revue de philosophie de la France et de l'étranger*, 491–512, special number on Brentano, ed. Guillaume Fréchette.
Mulligan, K. 2018a. How to Marry Phenomenology and Pragmatism. Scheler's Proposal (expanded version of a paper published in 2016). In *Pragmatism*

and the European Traditions: Encounters with Analytic Philosophy and Phenomenology Before the Great Divide, ed. M. Baghramian and S. Marchetti, 37–64. Routledge Studies in American Philosophy. London: Routledge.

Mulligan, K. 2018b. Logical Norms, Logical Truths & Normative Grounding. In *Bolzano and Grounding*, ed. B. Schnieder and St. Roski. Forthcoming.

Passmore, J. 1966. *A Hundred Years of Philosophy*. Harmondsworth: Penguin.

Perry, R.B. 1935. *The Thought and Character of William James. Vol. II, Philosophy and Psychology*. Boston and Toronto: Little, Brown and Company.

Picardi, E. 1991. Bemerkungen zur Auseinandersetzung zwischen Frege und Kerry. *Erlangen Historisches Colloquium*, Heft 4: 4–22.

Picardi, E. 1994. Kerry und Frege über Begriff und Gegenstand. *History and Philosophy of Logic*, XV: 9–32.

Russell, B. 1899 (Review of Meinong 1896). *Mind*, New Series, 8 (30, April): 251–256.

Russell, B. 1904. Meinong's Theory of Complexes and Assumptions. *Mind* 13: 204–219, 336–354, 509–524. Reprinted in Russell 1973, 21–76.

Russell, B. 1905a. On Denoting. *Mind* 14: 479–493. Reprinted in Russell 1956, 41–56, in Russell 1973, 103–119, and in Russell 1994, 415–427.

Russell, B. 1905b (Review of A. Meinong). Untersuchungen zur Gegenstandstheorie und Psychologie. *Mind* 14: 530–538. Reprinted in Russell 1973, 77–88.

Russell, B. 1906 (Review of Meinong 1906). *Mind*, New Series, 15 (59, July): 412–415.

Russell, B. 1906–1907. On the Nature of Truth. *Proceedings of the Aristotelian Society*, New Series, 7: 8–49.

Russell, B. 1907 (Review of Meinong 1907). *Mind*, New Series, 16 (63, July): 436–439.

Russell, B. 1911. Le réalisme analytique. *Bulletin de la société française de philosophie*, 11: 53–61.

Russell, B. 1973. *Essays in Analysis*, ed. D. Lackey. London: Allen & Unwin.

Russell, B. 1979 (1903). *The Principles of Mathematics*. London: George Allen & Unwin.

Russell, B. 1986 (1912). *The Problems of Philosophy*. Oxford: Oxford University Press.

Stout, G.F. 1896. *Analytic Psychology*, Vol. I, II. London: Swan Sonnenschein.

Stout, G.F. 1899. *A Manual of Psychology*. London: W. B. Clive.

Stout, G.F. 1929. *A Manual of Psychology*, 4th ed. London: W. B. Clive.

van der Schaar, M. 2013 (1991). *G. F. Stout and the Psychological Origins of Analytic Philosophy*. London: Palgrave Macmillan.

van der Schaar, M. (2017). Brentano, Moore, and Stout. In *The Routledge Handbook of Franz Brentano and the Brentano School*, ed. U. Kriegel, 349–357. Routledge.

Truth, Ascriptions of Truth, and Grounds of Truth Ascriptions

Reflections on Bolzano and Frege

Wolfgang Künne

In Sect. 1 of this chapter, I shall discuss Bolzano's attempt to give a definition of the concept of truth, in Sect. 2 I shall ask whether Frege succeeds in showing that all such endeavours are doomed to failure. In this chapter I shall remain neutral as to the question of definability, but

Both Frege and one of Europe's most important Frege scholars read Bolzano at least once. The latter happened when Eva Picardi worked on her pioneering paper on Frege and Kerry (*History and Philosophy of Logic* 15, 1994, 9–32). It was Benno Kerry, not Husserl, who was the first among Brentano's students to study Bolzano's Opus magnum against their nominalistically minded master's will, and he quoted very extensively from Bolzano's *Wissenschaftslehre* in a long series of papers to one of which Frege replied in 'Ueber Begriff und Gegenstand'. So I am fairly certain that Eva would have been interested in seeing the grandfather of analytical philosophy (*1848) confronted with its great-grandfather (†1848), as her teacher and fatherly friend Michael Dummett called these gentlemen while she was listening. How sad that at this point I had to use the phrase 'would have been'.

W. Künne (✉)
University of Hamburg, Hamburg, Germany
e-mail: wolfgang.kuenne@uni-hamburg.de

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_3

I will be much concerned with a Fregean thesis that apparently supports the key premise of his alleged proof of indefinability. The equivalence schema 'The thought that things are thus and so is true if, and only if, things are that way' captures an important feature of the concept of truth. Frege went beyond this *true-iff* principle when he claimed that the two halves of such biconditionals do not only stand and fall together,—they even express one and the same thought. It is doubtful whether Frege has any good argument for this Identity Thesis. In Sect. 3 of this chapter I will give reasons for this doubt. In Sect. 4 I shall show that, and why, Bolzano rejects the Identity Thesis. Bolzano emphasizes an important feature of our concept of truth that is not captured by the equivalence schema. One can hint at this additional feature by saying, 'If the thought that things are thus and so is true, then it is true because ofthings' being that way, and not *vice versa*'. In Sect. 5 I shall locate this *true-because-of* principle in the theory of grounding (*Abfolge*) that Bolzano outlined in the second volume of his monumental *Wissenschaftslehre* (henceforth: *WL*). In Sect. 6 I shall explore whether the Identity Thesis can be refuted by appealing to (the Bolzanian reformulation of) the true-because-of principle. On the following pages, I shall not try to argue for the true-because-of principle. Like Aristotle and Bolzano I shall accept it as a basic intuition concerning truth.[1] The brief Appendix points to a use of the notion of grounding that has been neglected in recent literature although Bolzano deemed it to be of great importance.

1 Bolzano's Definition of *Truth*

The proposition (*Satz an sich*) that is expressed by (co-referential utterances of) the sentences 'Socrates is courageous', '*Socrate è coraggioso*', '*Sokrates ist mutig*' and many others is true just in case Socrates has courage of one kind or another. This is the intuition that stands behind Bolzano's attempt at giving a decomposing definition of the concept of truth.[2] This definition runs as follows:

> (Df. T) A proposition *x* is true if, and only if, every object that falls under the subject concept of *x* has a property that falls under the predicate concept of *x*. [*Ein Satz an sich ist genau dann wahr,*] *wenn jeder Gegenstand, der der Subjectvorstellung des Satzes untersteht, eine Beschaffenheit hat, die der Prädicatvorstellung desselben untersteht.*[3]

A proposition that is not true is false. In Bolzano's (as in Aristotle's) eyes the quantifier 'every' has existential import, so according to (Df. T) there is

at least one object that falls under the subject concept of a true proposition. Sometimes exactly one object falls under the subject concept of a truth, sometimes several. Hence the definiens of (Df. T) is just as applicable to the proposition that all assassinators of tyrants have courage, as it is applicable to the proposition that Socrates has courage.[4] In the Philosophical Grammar that occupies large stretches of volume II of his *WL*, Bolzano tries very hard to support his conjecture that each and every proposition can be expressed by a sentence instantiating the schema 'A has b' (where 'A' is taken to be a dummy both for singular terms and for phrases of the form 'every such-and-such' and 'b' is a placeholder for singular terms that denote properties).

If (Df. T) is correct then as regards generality the predicate concept in a truth is in the same boat as the subject concept: in some truths exactly one property falls under the predicate concept, in some truths several properties do.[5] If the second state of affairs obtains, in Bolzano's words: 'if the predicate concept [of a proposition *x*] represents several properties'[6] then *x* is true, according to (Df. T) just in case every object that falls under the subject concept of *x* has at least one of those properties. In section 131 of *WL* Bolzano poses the question 'whether when we ascribe the property b to an object A we also ascribe to A all properties that fall under [the concept] *b*',[7] and he declares it to be obvious that the answer is negative. In order to make this obvious to his readers he employs this example:

(S1) Caius has intelligence.

In assertively uttering (S1) we surely do not commit ourselves to the claim that the Caius we are talking about has 'every kind of intelligence one can think of, for example, a well-developed, as well as a crude, a human, and an angelic intelligence, etc.' (loc. cit.). If (Df. T) is correct then the proposition expressed by an utterance of (S1) is true if, and only if, Caius has at least one of the many kinds of intelligence that (Bolzano takes to) fall under the concept of intelligence. Similarly, there are several kinds of courage, such as civil courage and military courage,[8] and in stating that Socrates has courage we do not commit ourselves to the claim that Socrates has every kind of courage. According to (Df. T), the proposition that Socrates has courage is true just in case Socrates has at least one of the kinds of courage that (allegedly) fall under the concept of courage.

But is the relation between the many kinds of courage and the concept of courage really that of falling under? What falls under the concept *courage*, I dare say, is the property of being courageous—and nothing else. The phrase '***a*** property of being courageous' does not even make sense.

Of course, this uniqueness claim does not exclude that there are different kinds of courage. But they do not fall under the concept *courage* but rather under the concept *a kind of courage*. Civil courage, for example, falls under this concept in the same way in which Socrates falls under the concept *a human being*. But civil courage does *not* stand in this relation to the concept *courage* any more than the kind Indigenous Australian falls under the concept *a human being*. Therefore, I propose to modify (Df. T):

> (Df. T)* A proposition *x* is true if, and only if, every object that falls under the subject concept of *x* has *the* property that falls under the predicate concept of *x*.

Even after this revision of (Df. T) section 131 of *WL* contains an important message: If a property b is such that there a several kinds of b, then the proposition that (every) A has b is true just in case (every) A has some kind of b, and the proposition leaves open *which* kind of b it is that makes the proposition true.[9]

Of course, this revision of (Df. T) does not address the obvious question that Bolzano's decomposing definition of the concept of truth (just like Aristotle's) has to face: is it really broad enough to cover all-truth candidates? Can every truth-candidate be expressed by an ascription of a property to one or more objects? Take any utterance of

> (S2) As it was rather cold in the morning, it may have been snowing for many hours.

Which property could be said to be ascribed in this utterance to which object(s)? It would not help to solve this Procrustes Problem[10] if one were to resort to a Tarskian recursive strategy and to take something like 'A non-compound non-quantified one-place predication expresses a true proposition if, and only if, the object that is referred to in that predication has the property that is predicated of it' as base clause. After all, (S2) contains two non-extensional operators ('as', 'may'), a feature-placing construction with 'it', tenses, and a non-standard quantifier, and none of this is provided for by Tarski's machinery.

2 Frege on the Indefinability of *Truth*

In his criticism of all attempts to define the concept of truth Frege begins with a 'local' argument against a definition of truth as 'correspondence with something real (*Übereinstim-mung mit etwas*

Wirklichem)'.[11] His example for something real is not a fact but a thing, the Cologne Cathedral.[12] According to Frege, a fact is nothing but a true proposition (*Gedanke*), and he rightly assumes that the correspondence meant in the definition is a relation of a truth bearer to something else.[13] I shall put his local argument aside, for it raises a host of exegetical issues,[14] and with respect to Bolzano's (Df. T) it is toothless anyway. Bolzano carefully avoids the predicate '*x* corresponds with (*stimmt überein mit*) *y*'.[15] He complains about its 'shifting meaning (*schwankende Bedeutung*)' in traditional accounts of truth, and he especially disapproves of the connotation of similarity (that plays a key role in Frege's local argument). So let us focus on Frege's 'global' argument:

> Every … attempt to define truth … collapses. For [1] in a definition marks (*Merkmale*) would have to be specified. And [2] in application (*Anwendung*) to any particular case the question would always arise whether it were *true* that the marks were present. So [3] one would go round in a circle (*So drehte man sich im Kreise*). Consequently, [4] it is probable (*wahrscheinlich*) that the content of the word 'true' is sui generis (*ganz einzigartig*) and indefinable.[16]

Surprisingly many commentators have misconstrued this objection as a vicious infinite-regress argument although [3] makes it as clear as can be that Frege takes himself to be offering a kind of vicious-circle objection. This does not mean that he condemns all definitions of 'true' as circular definitions. A definition is circular if the concept expressed by the definiendum is (openly or covertly) expressed by a proper part of the alleged definiens. Thus '*x* rotates = df. *x* moves around the axis of *x*' is a covertly circular definition, for what does 'the axis of *x*' mean if not: the line around which *x* rotates or might rotate? The circle Frege is complaining about is not of this kind. According to [2], it is a circle one gets into as soon as one tries to *apply* the alleged definiens of 'true'.

Unfortunately, as it stands the argument is enthymematic. So one must try to find the missing premise(s) acceptance of which can reasonably be ascribed to Frege. The definitions referred to in [1] are not what Frege calls constructive (*aufbauend*) or stipulative definitions. Such a definition either introduces a new expression for purposes of abbreviation (such as 'pi-meson' in physics) or forces an old expression into a new tightly circumscribed service (e.g. 'model' in mathematics). No philosopher who tries to define 'true' is aiming to graft a new meaning

upon this old word. The definitions at issue are rather dissecting (*zerlegend*) or analytic definitions which purport to capture, by means of a compound expression, the sense (or content, as Frege puts it [4]) of an atomic expression already in use.[17]

Now according to Frege's *technical* use of the term '*Merkmal*', M is a mark of the concept expressed by a predicate *P* just in case the following conditions are satisfied: M is a concept, M is expressed by a component of an analytic definition of *P*, and nothing can fall under the concept expressed by *P* unless it falls under M.[18] This condition is met if, and only if, the definiens of *P* is conjunctive. Thus, in virtue of the definition '*x* is a drake = df. *x* is a duck, and *x* is male', the concepts expressed by 'male' and by 'duck' are marks of the concept expressed by 'drake'. Such a definition pays epistemic dividends, for it allows us to answer the question whether the definiendum applies to an object *via* answering two simpler questions which do not presuppose that we have already mastered the definiendum. In this technical acceptation of the term 'mark' the concepts *brother* and *sister* are not marks of the concept *sibling* even though '*x* is a sibling = df. *x* is a brother, or *x* is a sister' does well as a definition, but here, too, we can answer the question whether the definiendum applies to an object by answering two conceptual-ly less demanding questions. Now Frege cannot reasonably assume that each and every candidate for a definiens of 'true' has the structure of a conjunction. The very definition he had attacked in his local argument lacks this kind of structure, and the same holds for (Df. T)*. So Frege's use of the term 'mark' in [1] and [2] cannot be the technical one of his earlier work. It seems to amount here to 'predicable concept that is a constituent of a complex concept'.[19]

But according to Frege one feature of conjunctive analytic definitions is shared by all adequate analytic definitions: they capture an order of epistemic priority,—we come to know that the complex concept expressed by the definiendum applies to an object *via* coming to know that a predicable constituent concept specified in the definiens applies to it. Consequently, a candidate for the title of an analytic definition of a predicate P is to be rejected if having the concept expressed by *P* is a precondition for deciding whether an alleged constituent concept applies in any given case. If we now assume, as Frege does in [2], that one cannot decide whether an alleged constituent of the concept of truth applies to a truth candidate without deciding whether it is *true* that it applies to

that candidate, the conclusion in [3] follows, for one can hardly decide whether something is true without having the concept of being true.

In the case of applying the Bolzanian (Df. T)* to a given proposition p_1, Frege could argue: According to its definiens of we have to decide whether p_1 has (1) the property of being such that at least one object falls under its subject-concept and whether p_1 has (2) the property of being such that an object has the property which falls under the predicate-concept of p_1 if it falls under the subject-concept of p_1. But—Frege would continue—one cannot decide whether p_1 has (1) and whether it has (2) without deciding whether it is *true* that it has either. So we would go round in a circle.

Frege's argument for indefinability crucially depends on the assumption that one cannot decide whether a concept that is said to be contained in the concept of truth holds of a truth candidate without deciding whether it is *true* that it holds of it. This assumption is an instance of a more general principle which Frege seems to regard as obvious:

> (Decide) One cannot decide whether *p* without deciding whether it is *true* that *p*.

In the next section, we will see why Frege finds this obvious: it follows from what I shall call his Identity Thesis.—It is noteworthy, though seldom noted, that his use of 'it is probable that' in [4] betokens that he himself does not regard his global argument as a watertight proof of his indefinability claim.[20]

3 Frege's Identity Thesis

In his paper 'On Sense and Reference' Frege famously said of instances of the schema 'The thought that *p* is true' that they express the same proposition as the corresponding less verbose instances of the schema '*p*':

> (Identity)$_1$
> One can … say: 'The thought that 5 is a prime number is true'. But on closer examination one notices that nothing more has been said than in the simple sentence '5 is a prime number'. (1892, 34)

In his paper 'The Thought' he makes the same identity claim about corresponding instances of the schemata 'It is true that *p*' and '*p*':

> (IDENTITY)$_2$
> The sentence 'I smell the scent of violets' has just the same content as the sentence 'It is true that I smell the scent of violets'. (1918, 61; cp. 1897, 153//141)

So, according to Frege, no matter whether a sentence is surrounded by the truth frame 'The thought that (…) is true' as in IDENTITY$_1$ or preceded by the truth prologue 'It is true that (…)' as in IDENTITY$_2$, the propositional content of the whole sentence coincides with that of the embedded sentence. The presence or absence of a truth operator of either kind makes no difference to the content of a declarative sentence as uttered in a certain context. Or so Frege thought.

Frege's Identity Thesis, if correct, justifies his acceptance of the principle DECIDE that we saw to underlie his indefinability argument. The universal closure of 'One cannot decide whether *p* without deciding whether it is *true* that *p*' is weaker than the universal closure of 'The proposition that *p* is identical with the proposition that it is true that *p*'. After all, the truth that one cannot decide whether *one's glass is half-full* without deciding whether *one's glass is half-empty* does not entail that the embedded sentences express one and the same proposition. But if the Identity Thesis is correct then Frege rightly endorsed the principle DECIDE.

One would like to see Frege actually perform the 'close examination' mentioned in IDENTITY$_1$ that allegedly reveals the correctness of this claim. He complies with this request when he argues for the identity thesis from (what I propose to call) the *illocutionary redundancy* of truth operators. In 'The Thought' Frege shows that the presence of this operator does not guarantee that an utterance has assertoric force, and that its absence does not put assertoric force at risk. By prefixing the truth operator to a sentence one does not ensure that an utterance of that sentence is an assertion: The antecedent in (an utterance of the form) 'If it is true that *p* then *q*' has no more assertoric force than has its truth-free counterpart, and if the actor who plays Marcellus were to surprise the audience by saying, 'It is true that something is rotten in the state of Denmark', his utterance would be just as much a mere *as-if* assertion as the utterance of an actor who shows more respect for Shakespeare's

words. On the other hand, if a speaker who begins his assertive utterance with the truth prologue had supressed the prologue he would not thereby have failed to make an assertion. Now Frege goes on to say: 'This explains why' presence or absence of a truth prologue or a truth frame has no effect on the propositional content of an utterance.[21] So he argues from *illocutionary* redundancy to *propositional* redundancy. But does the illocutionary-redundancy claim really lend any support to the propositional-redundancy claim? Can one not coherently endorse Frege's eminently plausible contention that presence or absence of a truth-prologue or a truth frame has no effect on the illocutionary force of an utterance, and reject IDENTITY$_{1-2}$?

Corresponding instances of 'It is true that *p*' and '*p*' are not only intensionally equivalent (the propositions they express have the same truth-value in every possible world), they are even cognitively equivalent in the following sense: two sentences are COGNITIVELY EQUIVALENT if, and only if, nobody who fully understands both can rationally assent to one of them without immediately being ready to assent to the other one as well. At several places Frege appealed to cognitive equivalence thus understood as a *necessary* condition of propositional identity but he never invoked it as a sufficient condition, and for various reasons that was wise of him. At one point in his *Posthumous Writings* he offers a sufficient condition for propositional identity that *invokes* the concept of cognitive equivalence[22]:

> (SUFFICIENCY)
> Two sentences express the same proposition if (1) they are cognitively equivalent and (2) neither of them is, or contains, a part which is such that one cannot fully understand it without immediately being ready to assent to it – or without immediately being ready to dissent from it.

If SUFFICIENCY were acceptable it would justify the restricted claim that the result of applying a truth operator to a sentence S has the same content as the plain sentence S *provided that* S does not defy dissent. Hence not much is to be gained by appealing to SUFFICIENCY, for clearly Frege's Identity Thesis is supposed to hold across the board. Worse than that, the criterion is not acceptable anyway. Consider the pair 'On May 20, 2020, a teacher will point twice at a *geometric figure* that is square' and 'On May 20, 2020, a teacher will point twice at a *rectangle* that is square'. Each of these sentences is such that one can very well

understand it without being ready to assent to it, and they are cognitively equivalent. But the place at which they differ is occupied by predicates that do not even have the same extension, so how could these sentences have the same propositional content?[23] After all, Frege maintains in complete generality the following principle of composition:

> (Composition)
> The sense of a part of a sentence is part of the sense of the sentence, that is, of the thought expressed by that sentence. (1919b, 156//98)[24]

Furthermore, if Sufficciency were correct then

(a) 5 is prime

would come out as identical in content not only with (b) and (c),

(b) It is true that 5 is prime
(c) The proposition that 5 is prime is true,

but also with

(d) Anyone who were to believe that 5 is prime would be right in so believing.

But the sense of the elaborate belief frame that surrounds (a) in (d) is glaringly absent from the sense of the arithmetical sentence (a). At this point, one is tempted to assert that the same holds of the sense of the truth prologue in (b) and of the truth frame in (c). Frege would not yield to this temptation, to be sure, but his line of resistance is baffling:

> (Self-effacement)
> The word 'true' has a sense that contributes nothing to the sense of the whole sentence in which it occurs as a predicate. (1915, 272//252)

For one thing, this is somewhat carelessly formulated. It is not the word 'true' that is a predicate in Frege's acceptation of this term, but rather the phrase '(…) is true'. More importantly, if we remove the *predicate* from a sentence that can be used to assert something, the remainder—just by itself, without any help from the linguistic context—can no

longer serve this purpose. What remains of (c) when the predicate '(…) is true' is removed can no longer be used to assert anything, let alone to assert that 5 is prime. And the same holds for (b) if we assume (with Bolzano) that this sentence, too, contains that predicate. (I shall return to this assumption.) By itself, the *designator* of a proposition can no more serve the purpose of assertion than can the designator of a number. So, let us reformulate Self-effacement in such a manner that this problem disappears:

(Self-effacement)*
A truth operator has a sense that contributes nothing to the sense of the sentence whose prologue or frame it is.

As for this contention, I must confess that I am simply mystified by it. How could *any* expression that has a sense fail to contribute it to the sense of a sentence that contains it? How could its presence or absence fail to make a difference to the proposition that is expressed?

4 Bolzano's Difference Thesis

As we saw in Sect. 1, Bolzano—like Frege—takes propositions to be structured entities that consist at bottom of nothing but sub-propositional senses. Like Frege, Bolzano embraces Composition, but, I am glad to say, he rejects Self-effacement. He maintains:

(Difference$_1$)
If *A is B* is a true proposition, then *the proposition that A is B is true* is undeniably also a true proposition, and in virtue of its composition the latter is a proposition that is different from *A is B. Wenn der Satz* [sc. *an sich*]*: A ist B, wahr ist: so ist unläugbar auch* [*der Satz:*] *der Satz, daßA B sey, ist wahr, ein wahrer Satz; und dieser ist seinen Bestandtheilen nach schon ein anderer, als der Satz* [sc. *an sich*]: *A ist B, selbst.* (*WL I*, 147)

The schematic truth ascription that Bolzano uses in Difference$_1$ has the same structure as the truth ascription Frege uses in Identity$_1$: the that-clause is preceded by a noun phrase. Borrowing a term from Jeffrey King, let us call expressions of the type 'the proposition that things are thus and so' *proposition-descriptions*.[25] As against Frege, Bolzano contends—and I wholeheartedly agree—that in a truth ascription containing

a proposition-description *two* sub-propositional senses are present, sc. the concept of a proposition and the concept of truth, that either do not occur at all in the proposition to which truth is ascribed or occur there less often. We must allow for the 'less often' possibility, since a truth operator can be applied to a sentence that itself contains a truth operator.[26]

In his posthumously published 'Paradoxes of the Infinite' Bolzano formulates his denial of the Fregean Identity Thesis as follows:

> (Difference$_2$)
> We find that the proposition that is expressed by the sentence 'A is true' is different from the proposition A, since it obviously has a very different subject…, for its subject is the whole proposition A itself. [*Wir finden*], *daß der Satz* [*an sich*], *welchen die Worte 'A ist wahr' ausdrücken, ein von A selbst verschiedener sei; denn* [*er*] *hat offenbar ein ganz anderes Subjekt… Sein Subjekt ist nämlich der ganze Satz* [*an sich*] *A selbst.* (1851, §13)

The subject of the proposition expressed by a certain sentence is whatever is designated by the subject term of that sentence.—How is the schematic letter 'A' used in the schemata Bolzano employs in Difference$_2$? It is a placeholder for a term that singles out a proposition and that can be substituted both in the *singular-term* schema 'the proposition A' and in the *sentence* schema 'A is true', hence it is a dummy for a *naked that-clause*. Hence sentences like 'That 5 is a prime number is true' are substitution instance of the sentence schema.[27]

Taking Difference$_1$ and Difference$_2$ together we see that Bolzano reckons with (at least) two styles of ascribing truth. The first is exemplified by instances of schema (s1), the second by instances of schema (s2):

(s1) The proposition that *p* is true.

(s2) That *p* is true.

So both a proposition-description and a naked that-clause single out what is declared to be true. Bolzano need not deny that both types of truth ascriptions can be given a truth-operator parsing, in other words, that both can be regarded as produced by inserting a sentence into a truth frame. But that parsing—he would insist—does not mirror the structure of the propositions that are expressed. Those propositions

have a subject concept under which a proposition falls, and this concept is either expressed by a naked that-clause or by a proposition-description.—At this point one wonders what Bolzano makes of instances of the schema

(s3) It is true that *p*.

Here is his answer: 'Sometimes the word "it" seems to be entirely superfluous, as in "It is true that etc."'[28] Bolzano regards the first word in (s3) to be what grammarians call a pleonastic pronoun or an expletive. (Italians renounce the luxury of an expletive: '*È vero che 5 è uno numero primo.*) Now in (s3) we cannot simply delete the expletive, we also have to change the word order. (Compare the Irish Regiment's song 'It's a long way to Tipperary'. Here, too, deletion of the expletive requires re-ordering: 'The way to Tipperary is long'.) It turns out that by Bolzano's lights (s2) and (s3) are just stylistic variants of each other. Once again, he can concede that a truth operator parsing of (s3) is permissible, that is, instances of (s3) can be constructed by prefixing a truth prologue to a sentence. But that parsing does not capture the structure of the proposition that is expressed. Now none of this is uncontroversial, but let us assume that Bolzano is right on each of these points.[29] Then instances of (s1), (s2) and (s3) have the same *predicational* structure as the next three sentences that *resist* a truth-operator parsing for the obvious reason that they do not contain any such operator:

(e) That is true.
(f) What the witness said is true.
(g) Goldbach's Conjecture is true.

Bolzano's unified predicational picture allows us to say, what seems to be intuitively correct, that in (g) as well as in

(h) It is true that every even number greater than 2 is the sum of two primes.

truth is ascribed to one and the same thing, namely to the proposition that every even number greater than 2 is the sum of two primes, to a proposition that unlike most other members of the family was deemed worthy of a proper name.[30]

5 Beyond T-Equivalence: Truth Ascriptions in Bolzano's Theory of Grounding

Bolzano emphasizes an important feature of the concept of truth that is not captured by the biconditional schema

(T-Equivalence)
It is true that *p* if, and only if, *p*.

Here are five ways of hinting at this feature that was first noted by Aristotle. Suppose the proposition that things are thus and so is true. Then [1] its truth is *due to* things' being thus and so, [2] it *owes* its truth *to* things' being thus and so, [3] it is true *thanks to*, [4] *in virtue of*, [5] *because of* things' being thus and so,—and not vice versa. I shall employ the last of these formulations, and I call the truth about being true they all try to capture the *true-because-of* principle. There are three passages in Aristotle to which this principle owes the nickname 'Aristotle's Insight' that it received in recent literature[31]:

Aristotle-1
It is not because of our having the true belief that you are pale, that you are pale; rather it is because of your being pale (*διὰ τὸσὲ εἶναιλευκόν*) that we who say so speak truly. (*Met.* Θ 10, 1051 b 6-9)
Aristotle-2
It is in virtue of the thing's being (*τῷτὸπρᾶγμαεἶναι*) … that the statement (*λόγος*) [that the thing is] is said to be true… (*Cat.* 5, 4 b 8-10)
Aristotle-3
Whereas the true statement (*λόγος*) is in no way the ground (*αἴτιος*) of the thing's being, the thing (*πρᾶγμα*) does seem in some way to be the ground (*αἴτιον*) of the statement's being true, for it is in virtue of the thing's being that the statement [that it is] is called true … (*Cat.* 12, 14 b 18–22)

In the main clause of the first sentence of (A-3) one would have expected 'the thing's being' rather than 'the thing', and both in (A-3) and (A-2) it is unclear whether the slippery word 'being' is used in the sense of 'existing' or as short for 'being thus and so'.[32] In any case, the example used in (A-1) shows that no restriction to existential statements is intended.[33] It should not go unmentioned that Aristotle also subscribed

to the *true-iff* principle,—in the fourteenth century Jean Buridan even came to call it *regula Aristotelis*.[34]

Quine registers the *true-because-of* feature of the concept of truth, but surprisingly he thinks that it *is* captured by rhetorically embellished instances of T-Equivalence: 'Truth should hinge on reality, and it does. No sentence is true but reality makes it so. The sentence "Snow is white" is true, as Tarski taught us, if and only if real snow is really white.'[35] Quine's suggestion that the feature in question is already taken into account by instances of T-Equivalence is clearly mistaken. The proposition that snow is white is true because of snow's being white, but not vice versa: it is not the case that snow is white because of this proposition's being true. The asymmetry of the because-of relation cannot be accommodated by means of a commutative connective even if we embellish the right-hand side of the Tarskian biconditional or its propositional variant by inserting the purely rhetorical ornaments 'real' and 'really'.[36]

Bolzano does better. He, too, embraces the *true-iff* principle:

> (T-Equivalence)$_{\mathbf{B}}$
> [There are] propositions that are inter-deducible, as, for example, the proposition A is true is deducible from A itself, and conversely. [*Es gibt*] *Sätze, welche sich wechselseitig auseinander ableiten lassen; wie etwa aus jedem gegebenen Satz A der Satz: A ist wahr, und wieder jener aus diesem ableitbar ist.* (*WL IV*, 414)

In the framework of Bolzano's theory of deducibility (*Ableitbarkeit*) this amounts to saying: every systematic exchange of concepts under which a given proposition comes out true is an exchange under which the corresponding truth ascription also comes out true, and vice versa.[37] But Bolzano clearly distinguishes T-Equivalence from the *true-because-of* principle that he also endorses:

> (T-Ground)
> Suppose that A is some truth. Then the truth that the proposition A is true is a genuine consequence of A. *Es sey … A was immer für eine Wahrheit: so ist die Wahrheit, daß der Satz A wahr sey, eine echte Folge aus ihr.* (*WL II*, 357)[38]

Don't be misled by the word 'consequence'. Under the Bolzanian reading of this word only truths have 'consequences', and 'consequences' are always truths:

> (Verity)
> Just as […] there are causes and effects only in the realm of actuality, so there are grounds and consequences only in the realm of truths. *Wie es … Ursachen und Wirkungen nur im Reiche der Wirklichkeit gibt: so gibt es Gründe und Folgen nur im Reiche der Wahrheiten*. (*WL II*, 356)

Verity is one of the principles of the tentative theory of 'grounding (*Abfolge*)', of the ground–consequence relation, that Bolzano sketches in volume II of his *WL*.[39] He distinguishes partial grounds (*Theilgründe*) from complete ones (*vollständige Gründe*)[40] and closest or immediate grounds (*nächste oder unmittelbare Gründe*) from more remote or mediate ones (*entferntere oder mittelbare Gründe*).[41] Let us briefly consider how these notions apply to truth ascriptions and their grounds.

Often the complete ground of a proposition is only provided by several propositions jointly. The complete ground of the truth that (Boscovich and Bolzano were priests), for example, is provided by a collection (*Inbegriff*) of truths that consists of the proposition that Boscovich was a priest and the proposition that Bolzano was a priest. Each truth in this collection is only a partial ground of the conjunctive truth. But when it comes to truth ascriptions, Bolzano argues, the situation is different. The proposition that ascribes truth to the true proposition A 'does not require for its grounding (*Begründung*) any further truth than A; hence A constitutes its complete ground' (*WL II*, 357). Of course, this does not imply that A itself does not have a ground. The complete ground of the proposition that it is true that Socrates or Seneca drank the hemlock is the truth that Socrates or Seneca drank the hemlock, and the complete ground of this disjunctive truth is the truth that Socrates drank the hemlock.

When I commented on Quine's failed attempt to capture the *true-because-of* feature of the concept of truth, I had not yet any reason to mark the difference between ascriptions of truth to propositions, as in Bolzano and Frege, and ascriptions of truth to sentences, as in the Tarskian paradigm Quine played with: 'Snow is white' is true if, and only if, snow is white.[42] Now we have reached a point at which there is a strong reason to underline that difference. The truth that snow is white is by no means

the complete ground of the truth that *the sentence* '*Snow is white*' is true. It needs help from a second truth, namely the proposition that 'Snow is white' means that snow is white. If that sentence were to mean that blood is red then the truth that snow is white would not even be a partial ground of the truth that the sentence 'Snow is white' is true.

Not only ascriptions of truth to a true proposition but also double negations of true propositions show that a truth may have just one ground. But no truth, Bolzano argues, has only one consequence (*WL II*, 358 ff.). The proposition that 5 is a prime number is the complete ground not only of the ascription of truth to that proposition but also of the proposition that it is not the case that 5 is not prime,[43] and of every proposition expressed by a disjunction in which '5 is a prime number' is the only disjunct that expresses a truth.[44] The case of truth ascriptions and the last two examples can be used to show that for *every* truth, regardless of its content, there is more than one truth of which it is the complete ground.[45]

As for the distinction between immediate and mediate grounds, look at the following series (Σ) of propositions where '*p*' serves as a stand-in for any sentence that expresses a truth:

(Σ) that *p*,
that it is true that *p*,
that it is true that it is true that *p*,
that it is true that it is true that it is true that *p*,
and so on *ad infinitum*.

In (Σ) each proposition is the immediate ground of its successor, and each proposition in (Σ) is a mediate ground of each of the infinitely many propositions that follow its successor. The relation '*x* is an immediate ground of *y*' is intransitive (*WL II*, 371), whereas the relation '*x* is a mediate ground of *y*' is transitive.

The series (Σ) deserves our attention for more than one reason. This series is one of the weapons Bolzano employs in the battle against global scepticism that he wages in the *Theory of Fundamentals* in his *WL*.[46] If the stairs of the staircase in my representation of (Σ) were occupied by different designators of the *same* proposition, as Frege would maintain, then one of Bolzano's proofs of the statement that there are infinitely many truths if there is one would fail badly (*WL I*, 147 n.). Furthermore, (Σ) shows that *some* non-conceptual truths satisfy a

requirement on grounding that Bolzano only imposes on a proper subset of truths, namely conceptual truths. The requirement I have in mind is this: If *x* is a (mediate or immediate, complete or partial) ground of *y*, then *x* must not be conceptually more complex than *y*. According to one of the criteria Bolzano employs at different places, *x* is a conceptually more complex proposition than *y* if, and only if, *x* contains not only all concepts of which *y* is composed but also at least one additional concept or at least one additional concept occurrence.[47] No matter whether the proposition that an instance of (Σ) starts with is the conceptual truth that five is prime or the non-conceptual truth that Clive is prim, no proposition in (Σ) that grounds some other proposition in (Σ) is conceptually more complex than the latter.

Another principle in Bolzano's theory of grounding justifies my contention that T-Ground accommodates the Aristotelian insight:

> (Asymmetry)
> The relation between ground and consequence can never be reciprocal. [*D*]*as Verhältniß zwischen Grund und Folge kann nie ein wechselseitiges seyn*. (*WL II*, 352)[48]

In virtue of this principle there is no need to add an 'and not vice versa' clause to T-Ground, and the explicit denial of reversibility in Aristotle's Insight was logically redundant.[49]

If grounding is an asymmetrical relation then it is also irreflexive. Hence the next principle can be classified as a theorem of the theory of grounding:

> (Irreflexivity)
> We can claim neither of an individual truth A, nor of a collection of truths A, B, C, D, … that they stand to themselves in the relation of grounding, i.e., that they are their own ground, their own consequence. *Wir können weder von einer einzelnen Wahrheit A, noch auch von einem Inbegriffe mehrer Wahrheiten A, B, C, D, … behaupten, sie ständen zu sich selbst in dem Verhältnisse einer Abfolge, d.h. sie wären der Grund und die Folge von sich selbst*. (*WL II*, 356)[50]

The principles Verity, Asymmetry and hence Irreflexivity hold for all grounds, no matter whether partial or complete, mediate or immediate.

Now Bolzano points out that a two-place predication of the form '*x* is a partial, or the complete, ground of *y*' expresses a truth just in case a compound sentence expresses a truth in which the connective 'because' is preceded by a sentence expressing *y* and succeeded by a sentence expressing *x* (*WL II*, 207, 221–222). Here a proviso has to be added: provided that the subordinate clause serves to explain *why things are* as the main clause says they are. (This contrasts with the evidential use of the connective 'because' when the subordinate clause is used to say *why it is reasonable to believe that things are* as the main clause says they are.[51]) Taking the exclusion of the evidential use of 'because' as understood one can say: the proposition that *q* is at least a partial ground of the proposition that *p* just in case (*p* because *q*). If we formulate the principles of grounding by means of this connective then Verity, Asymmetry and Irreflexivity get transformed into their '*connectival*' counterparts (henceforth marked by a superscript '©'):

(Verity)©	If (*p* because *q*) then (*p* and *q*)
(Asymmetry)©	If (*p* because *q*) then not (*q* because *p*)
(Irreflexivity)©	It is not the case that (*p* because *p*).

The preposition 'because of' in the English rendering of Aristotle's Insight I began with cannot stand in front of a sentence,—I've used gerund constructions like 'snow's being white' etc., where Aristotle's Greek requires infinitival constructions. If we replace 'because of' by the dyadic operator on *sentences* 'because', we need not look for objects of any kind to play the role of (what many philosophers are fond of calling) truth makers, for sentences are not names of components of reality. (That remains true even if we follow Frege in construing sentences, or rather, their *Begriffsschrift* counterparts, as names of truth-values.)

Strictly speaking, schemata are not principles, for they are not in the truth line of business. The principles in question are statements to the effect that all substitution instances of those schemata are true. Here the notion of a substitution instance requires some attention. Consider Irreflexivity©. Is it really the case that no substitution instance of the schema '*p* because *p*' expresses a truth? That very much depends on our standard for being a sentence of that form. If the syntactical congruence of the English sentence substituted for the first occurrence of '*p*' with the sentence substituted for its second occurrence suffices for being a compound sentence of that form, then such sentence may very

well express truths.[52] Suppose the word 'sick' is ambiguous in such a way that it can be used to ascribe a condition of body (being ill, for short) and a condition of mind (being depressed, for short). Then the sentence 'She is sick because she is sick' has a reading under which it expresses the proposition that she is depressed because she is ill, and that might very well be true. And, of course, when I first point to Claudia and then to her mother while saying, 'She is depressed because she is depressed', then I might say something true. The sentence 'Most people celebrate Christmas because must people celebrate Christmas' may have expressed a truth when Kurt Tucholsky, one of the most important journalists of the Weimar Republic, used it to say that most people meet the following condition: they celebrate Xmas because most people celebrate Xmas.[53] Such cases show that we require a more demanding conception of a substitution instance. A compound sentence is not a substitution instance of '*p* because *p*' unless the sentence-occurrences that flank the connective express (within this compound) one and the same proposition. This demand on substitution instances assorts well with the assumption that the truth-conditions of instances of '*p* because *q*' coincide with the truth-conditions of corresponding instances of 'The true *proposition* that *q* is a partial, or the complete, ground, of the true *proposition* that *p*'.

The principle T-Ground can be captured by three connectival formulations:

(T-Ground)© (1) If *p* then (the proposition that *p* is true because *p*).
(2) If *p* then (that *p* is true because *p*).
(3) If *p* then ((it is true that *p*) because *p*).

Henceforth I shall use (3) as the representative connectival rendering of the Aristotelico-Bolzanian insight because it is the most colloquial of the three. (Note that the inner pair of brackets in the consequent of (3) is required to forestall an unwanted reading.) A look at T-Ground© (3) suffices to see that the dyadic operator 'because' is hyper-intensional. Suppose 'p_1' is a sentence that expresses a truth. 'It is true that p_1' and 'p_1' are not only intensionally but even cognitively equivalent, and yet they cannot be exchanged *salva veritate* in the compound sentence 'It is true that p_1, because p_1'.

6 The *True-Because* Principle and the Identity Thesis

T-Ground© can be employed in a prima facie convincing argument against Frege's Identity Thesis. The argument I have in mind runs as follows:

(P1) If 5 is prime then ((it is true that 5 is prime) because 5 is prime).
(P2) 5 is prime. So,
(C1) It is true that 5 is prime, because 5 is prime.
(P3) The sentence in line (C1) is a 'because' sentence that expresses a truth.
(P4) A 'because' sentence never expresses a truth if the sentences that flank this connective express the same proposition. So,
(C2) 'It is true that 5 is prime' and '5 is prime' do not express the same proposition.

Obviously, the same kind of argument can be run for any pair of corresponding instances of 'It is true that *p*' and '*p*' that express truths. If the argument is valid and all its premises are correct then Frege is wrong in assuming that the members of such pairs always express the same proposition.

There is some reason to suspect that premise (P4) is the Achilles heel of the argument. Prima facie counter-examples are provided by statements that are backed by conceptual analysis.[54] Consider this statement:

(j) 5 is a prime number because 5 is a natural number greater than 1 that has no positive divisors other than 1 and itself.

Suppose we endorse (j) and maintain, as Bolzano certainly would, that the predicate applied to the number 5 in the subordinate clause and the predicate applied to it in the main clause have the same sense and hence express the same concept.[55] Then it has to be admitted that in (j) the same proposition is expressed twice, and yet (j) seems to express a truth. But if that appearance is not deceptive then the fourth premise of the above anti-Fregean argument is false. (Of course, if that argument vainly tries to refute the Identity Thesis that does not show that this thesis is correct.)

Even if the proposition that 5 is prime is twice-expressed in (j), it is obviously not expressed in the same way both in the subordinate

clause and in the main clause: in the latter, components of the concept of a prime number are separately expressed and thereby the structure of the proposition is revealed. At an important point in his philosophical grammar Bolzano explicitly acknowledges that one of two sentences expressing the same proposition can reveal more propositional structure than the other. Suppose two elementary predications differ only in one respect: one of them employs the construction *copula followed by a general term* (e.g. 'Socrates is courageous') while the other employs the construction *'has' followed by the nominalization of that general term* ('Socrates has courage'). Such sentences—Bolzano maintains—express the same proposition but—he continues—it is the second formulation that is 'more distinct (*deutlicher*)' (*WL II*, 10–11) because it 'makes the components of the proposition more salient (*läßt die Bestandtheile* [*des Satzes an sich*] *deutlicher hervortreten*)' (*WL II*, 47). This particular claim may be hard to defend, but there is no need to defend it here. What matters is only that Bolzano says about such pairs of sentences what he could say—with at least as much plausibility—about the clauses of (j).

Or consider this example. Intuitively, 'Nightingales exist (whereas centaurs don't)' expresses the same proposition as 'There are nightingales (whereas there are no centaurs)', but the latter formulation is more distinct: it sets 'Nightingales exist' far apart from 'Nightingales sing' in containing only one general term, thereby destroying the impression that existing is like singing something these birds do but unlike singing something they inaudibly do.

It is worth mentioning in this context that Frege also concedes that two sentences of the same language can express the same proposition in divergent ways. In 'Function and Concept' he says of two sentences that one of them 'expresses the same sense' as the other 'but in a different way (*drückt zwar denselben Sinn aus …, aber in anderer Weise*)'.[56] Of course, as it stands this is too unspecific. If different sentences express the same proposition then it is always the case that each of them does so in its own way, but no member of the trio 'Kaa is a snake', 'Kaa is a serpent' and '*Kaa ist eine Schlange*' makes the structure of the proposition expressed more salient than its partners do. That is a performance that distinct sentences can only deliver if they do not have the same logical form.

Not any old syntactical difference between two sentences makes for a difference of logical form. I take sameness of logical form to be relative to a logically regimented language and to a manual for translating

from an ordinary language L into a logically regimented part of L. Let us assume the regimented version of English contains only sentences that are syntactically congruent with substitution instances of schemata in classical first- and second-order quantificational logic with identity. Then 'Adèle and Barbara are French' and 'Adèle is French, and so is Barbara' do not differ in logical form from their canonical counterpart 'Adèle is French, and Barbara is French'.[57]

If premise (P4) of the anti-IDENTITY argument I constructed is false, then this reveals a problem for Bolzano's theory of grounding.[58] In his theory the relata of grounding are always propositions. But if the predicates in a true 'because' statement are the analysandum-expression and the analysans-expression of a conceptual analysis, then the relata of the grounding relation are more fine-grained: a proposition as articulated in one way and the same proposition as articulated in another way. These relata could be represented as pairs of a proposition and a schema. In the case of (j), the grounding relatum could be represented by a pair consisting of the proposition that 5 is prime and a schema of a complex conjunction while the ground*ed* relatum is represented by a pair consisting of the same proposition and the schema of an elementary monadic predication.

If we share Bolzano's conviction that the sentences that flank the 'because' in (j) and its ilk express one and the same proposition then there seems to be just one way in which Bolzano's assumption that the relata of the ground-consequence relation are always propositions can be defended against the *prima facie* counter-examples provided by such sentences. One must insist that *secunda facie* they turn out not to be counter-examples. If they are not counterexamples then appearances are deceptive: sentences like (j) only seem to express truths. This is actually the stance towards (j) I want to recommend.

Those who take (j) to be a literal expression of a truth are, I think, victims of a subtle confusion. Since the subordinate clause employs the analysans-expression of the analysis of the concept of a prime number, they understand (j) as if it were partly about a concept, that is, they misinterpret its main clause as if it had the same sense as the main clause of the following sentence that does literally express a truth:

(k) 5 falls under the concept of a prime number because
5 is a natural number greater than 1 that has no divisors other than 1 and itself.

The confusion may be fostered by the fact that both in the case of (j) and in the case of (k) the subordinate clause is not only intensionally but also cognitively equivalent with the main clause.[59]

The decisive difference between (j) and (k) is that the two sentences flanking 'because' in (k) do *not* express one and the same proposition, for the proposition expressed by the main clause contains the concept of *falling-under* and the concept of a *concept* that are both plainly absent from the purely mathematical proposition expressed by the subordinate clause. It is in virtue of this difference that the truth expressed by (k) does *not* falsify premiss (P4) of the anti-IDENTITY argument. If this way of treating sentences like (j) is correct, then Bolzano's principle T-GROUND really poses a serious threat to Frege's Identity Thesis.

People who assent to (j) need not regard it as literally expressing a truth. They can take (j) to convey only indirectly a true message, namely the truth that is literally expressed by (k). After all, even sentences that obviously express falsehoods when understood literally can be used to convey truths. 'Enough isn't enough' might be used to convey a truth to the effect that what certain people deem to be enough isn't really enough. That does not show that 'Enough isn't enough' falsifies the claim that all substitution-instances of 'What is F is not F' express falsehoods (if both occurrences of the predicate make the same contribution to the proposition), for what 'Enough isn't enough', literally understood, expresses is as false as can be. Bolzano describes the other side of the same coin when he writes: 'People say frequently, "What is bad is bad"; taken literally (*seinem Buchstaben nach*), this is indeed an empty tautology. Yet, what we actually think when using these words and what we also want to convey (*zu verstehen geben*) by means of them is presumably something quite different.'[60]

Would Bolzano be ready to adopt the stance towards to (j) and its ilk that I favour? For one thing, he would not dislike the fact that it allows him to stick to his claim that only propositions (and collections of propositions) are relata of the ground-consequence relation. But there is more than the defensive potential of this stance that supports the hypothesis that he would find it congenial. As I pointed out above, Bolzano believes that 'Socrates has courage' and 'Socrates is courageous' express the same proposition but that the first sentence 'makes the components of the proposition more salient'. But he never claimed Socrates is courageous because he has courage. I believe he never maintained this because he thought (and rightly so, I dare say) that it is false. Similarly, 'There are

nightingales' and 'Nightingales exist' express the same proposition but the first sentence makes the structure of that proposition more salient. But it is not the case that nightingales exist because there are nightingales. Now Bolzano could take the same stance towards the sentences combined in (j): one of them makes the structure of the proposition they both express more salient than the other, but the compound sentence (j) does *not* express a truth. This would not, and should not, prevent him from conceding that 'what we actually think in speaking these words and also want to convey' is 'something quite different'—namely the *truth* that is literally expressed by (k).[61]

Notes

1. Any attempt at a proof of this principle from a definition of truth presupposes, of course, that pace Frege such a definition is to be had.
2. In my flow chart of conceptions of truth in (2003) I have characterized Bolzano's view as a borderline case of a 'Classical, or Object-based, Correspondence' account of truth (as opposed to 'Cambridge, or Fact-based, Correspondence' accounts): op. cit. 3–5, cp. 94–112, esp. 108–110. I shall explain in Sect. 2 why I call his view (as well as Aristotle's) a borderline case.
3. Bolzano (1834b) 111, cp. (1834c) 105. I took the liberty of replacing Bolzano's metaphor 'stand under (*stehen unter*)' by the more familiar metaphor 'fall under' that has become the common coin ever since Frege used it.—Translations from German and Greek are my own. References to works by Frege or Bolzano are always by original pagination that can be found in the margins of modern editions and translations. Since Bolzano's works are even in Austria and Germany far less easily available than Frege's, I shall always append the German wording to my quotations from Bolzano.
4. And according to (Df. T), the propositions that all Amazons are courageous, and that Penthesilea is courageous are equally false.
5. Bolzano, *WL II*, 26 f. Some propositions of the form 'A has (property) b' are such that *no* property falls under the concept expressed by 'b'. Bolzano forgot to give us an example. Let me make good for the omission: (i) The proposition expressed by the next sentence contains a concept under which, as a matter of fact, no property falls: 'Frege has *the property that Bolzano ascribed to him on Good Friday 1830*.' (ii) The proposition expressed by 'Frege has *the property that Bolzano both had and lacked on Good Friday 1830 at high noon*' contains a concept under which no property *can* fall.

6. '*sofern die Prädicatvorstellung der Beschaffenheiten mehre vorstellt*' (*WL II*, 26).
7. '*ob wir, indem wir einem Gegenstand* A *die Beschaffenheit* b *beilegen, demselben auch alle der Vorstellung b unterstehende Beschaffenheiten beilegen*' (loc. cit.).
8. Cp. the sermon 'On Courage' that Bolzano delivered as university chaplain on 25.02.1810, in (BGA) 2A.17/1.
9. Cp. the exegetically more detailed discussion of (Df. T) in my (2015b), bes. 399–405.
10. See my (2003) 111 f, 126, 139 n., 317.
11. Frege (1918) 60.
12. Frege (1897) 140//129, (1918) 60. Obviously, the view Frege is about to attack is a paradigm case of an Object-based Correspondence account of truth.
13. Frege (1918) 74 and 60.
14. I try to cope with them in (2010a) 394–397.
15. Bolzano, *WL I*, 127–132, 179–180, 200 and (1834c) 128, 163.
16. Frege (1918) 60. Frege underwrites the conclusion of this argument already in (c. 1883) 189//174 and in (1897) 140//129. In my flow chart of conceptions of truth in (2003) I have characterized Frege's view as one version of alethic Primitivism: op. cit. 12–18, esp. 16, cp. 129–133.
17. Cp. Frege (1914) 224–229//207–211.
18. For references and critical discussion see my (2001) Sect. 3 and (2010a) 219–225. Frege introduced the technical term 'mark' at a time when he had not yet distinguished between *Sinn* and *Bedeutung*. That made for some tensions in his later employment of this term. I call here 'concept' what Frege came to call (not a concept but) the *sense* of a concept-word.
19. As for the requirement of predicability, consider '*x* is a bachelor = df. *x* is an unmarried man in marriageable age'. The concept of negation expressed by the prefix 'un-' in the definiens is not a predicable concept.
20. As some readers of this chapter might know, I side with Bolzano in taking the concept of truth to be definable, but as announced in the summary, on the following pages I shall remain neutral on this issue. I have first outlined my Modest, or Quantificational, Account of Truth in (2003) 333–374. It avoids the Procrustes Problems from which Bolzano's definition suffers, but it hardly needs saying that it has its own problems. I try to clarify some issues in (2008c), and I defend it against Boghossian's objections in (2010b).
21. Frege (1918) 63.
22. What follows is my own rendering of the claim Frege makes in (1915) 213//197.

23. The assumption behind my rhetorical question might provoke an *objection*: The sentence (A) 'Socrates defends Socrates' contains the predicate '(…) defends Socrates' where the sentence (B) 'Socrates is a self-defender' contains a predicate with a different extension, and yet (A) and (B) express the same proposition.—*Reply*: There is a good reason for rejecting the second conjunct. (A) and (B) are cognitively equivalent, to be sure, but as we know that doesn't guarantee propositional identity. (A) is conceptually less demanding than (B). In order to understand (A) you need only grasp the sense of 'Socrates' and that of a predicate that also occurs in 'Cicero defends Murena'. In order to do that you need not grasp the sense of '( ) is a self-defender'. Cp. Dummett (1981) 341, (1991) 173 ff, (1997) 247 f. (Dummett's argumentation is slightly marred by the false assumption that '(…) kills himself' has the same sense as '(…) commits suicide'. Firstly, not only male persons commit suicide. Secondly, a man might kill himself without intentionally doing so, and then he does not commit suicide.)
24. Composition occurs *verbatim* also in Frege (1913) 20. Frege embraced this principle already in volume I of his *Grundgesetze der Arithmetik*: "If a name is part of the name of a truth-value then the sense of the former is part of the thought expressed by the latter" (Frege 1893, §32). In *Grundgesetze* expressions of a thought are taken to be names of truth-values, and each sense-carrying component of a name of a truth-value is also classified as a name.
25. King (2007) 137.
26. Talk of one and the same abstract entity occurring several times in another abstract entity is not unproblematic. For a discussion see Künne (2001) Sect. 4, 'The Repetition Problem'.
27. While the sentence schema 'A is true' is equally hospitable to 'Goldbach's Conjecture', 'Platonism', 'that snow is white', 'the proposition that snow is white', 'what the witness said' and 'that', the singular-term schema 'the proposition A' permits only insertion of the third of these expressions (and its ilk).
28. *Zuweilen scheint das* Es *ganz überflüßig da zu stehen, wie in dem Ausdrucke: Es ist wahr, daß u.s.w.* (Bolzano, *WL II*, 216). Bolzano goes on to say: "For after all, this is completely equivalent [*durchaus gleichgeltend*] with: 'The proposition that etc. has truth'." It can hardly be his considered view that the sense of 'It is true that snow is white', for example, contains the concept of a proposition. (As for his replacing 'is true' by 'has truth', see below Sect. 6.) In the statement just quoted he takes '*x* and *y* are completely equivalent' to amount to the same thing as '*x* and *y* have the same sense', for he continues: "Similarly, the idiom: It's fine

weather today [*Es ist heute schönes Wetter*] means the same as [*eben so viel heißt als*]: Today's weather is fine."

29. Arguments for this assumption can be found in Künne (2013, 2015a).
30. I do not regard the subject term of (g) as a definite description because unlike its *small-c* variant it would single out the proposition that every even number etc. even if Goldbach never conjectured anything, let alone that every even number etc. (Capitalization can lead to semantic 'devaluation': the Holy Roman Empire of the German Nation lacked various properties that the adjectives in its title signify.) One might still doubt that 'Goldbach's Conjecture' is a proper name. (In Strawson (1950) 21, 24 such phrases are classified as '"impure" proper names' and 'quasi-names'.) If you take 'Platonism' to be a clearer example of a proposition name, replace (g) by 'Platonism is true' and (h) by 'It is true that there are abstract objects'.
31. This nickname as used in Künne (2003) has become the common coin in the pertinent literature. Yet it seems to have fallen into oblivion that it was Chisholm who introduced it in his (1960) 113. As was to be expected, the Aristotelian insight was not lost upon Aquinas (c. 1271) §1897: '*Dispositio rei est causa veritatis in opinione et oratione*. (The state of the thing is the ground of truth in belief and speech.)'
32. Austin is reported to have said: 'When God called out to Moses from the burning bush, "I am", the only proper reply for Moses to have made was "You are what?".'
33. For some reflections on exegetical problems and a discussion of Aristotle's account of truth see my (2003) 95–101 and 150.
34. Aristotle formulates it in the lines that precede (A-3), i.e. in *Cat.* 12, 14 b 15–18, and in *De Int.* 9, 18 a 39–42. In fact, he uses instances of the *true-iff* principle, but he clearly intends a general message. Cp. Buridanus (c. 1350) chapter VIII, 2nd Sophisma, 45, 47.
35. Quine (1970) 10. The fact that Tarski and Quine (unlike Bolzano and Frege) take sentences to be truth-value bearers makes no difference to the point I am currently driving at. But we will soon reach a point where this difference really makes a difference.
36. Dummett (1959) 14 clearly avoids Quine's mistake of blurring the distinction between the *true-iff* principle and the *true-because-of* principle. Cp. Künne (2007) 315–320 and Dummett's Reply (2007) 345–347.
37. For a concise and precise reconstruction of Bolzano's theory of deducibility (with references to the literature) see Morscher (2014).
38. Cp. Bolzano, *WL II*, 370, 374. In T-Ground I again took the liberty of deleting misleading quotation-marks. Was Bolzano inspired by Aristotle when putting forward this principle? He never refers to A-1 or A-2. The suggestion to the contrary on p. 11 of Tatzel (2002), a pioneering study

on Bolzano's theory of *Abfolge*, is rightly criticized in the first book-length discussion of this theory: Roski (2017) 71 n. 56. The reference in Tatzel's footnote 42 is not really pertinent. Bolzano does not refer to A-3 either, but in *WL II*, 364 we see that he read Chapter 12 of *Categories* that contains A-3. This can be seen as (admittedly weak) evidence for the conjecture that he was inspired by Aristotle when he formulated T-Ground even though at that point he did not bow to Aristotle.

39. Bolzano, *WL II*, 191–194, 207, and especially 339–390. I use 'principle' as an abbreviation of 'axiom or theorem'. Verity has the status of an axiom in Bolzano's theory, and rightly so. In order to shelter off false associations, I prefer to render '*Abfolge*', as has become common in the literature, by the somewhat stilted term 'grounding' (rather than by 'entailment'). Bolzano's own terminological policy at this point is sub-optimal. '*Abfolge*' is at home in talking about events: the *Abfolge* (sequence) of certain events may be such that they take place *in rascher Abfolge* (in quick succession). An appropriate label for what *Gründe* are grounds of is not far to seek: '*Begrün-detes*'. Bolzano himself quotes Latin philosophical texts that use the analogous pair '*ratio*' and '*rationatum*' (*WL II*, 372, 560). A natural title for an argument that gives *Gründe* for its conclusion is '*Begründung*', and the phrase '*objektive Begründung*' would be an appropriate designation for the relation between truths that are *Gründe* and *Begründetes*. I shall soon quote a passage where Bolzano himself uses *Begründung* for this purpose.
40. Bolzano, *WL II*, 207, 340–341.
41. Bolzano, *WL II*, 377–378, 389.
42. If you think that it is simply wrong to ascribe truth to sentences, take 'true' to be short for 'expresses a truth'. Then the contrast is that between ascribing the property of being true to a proposition and ascribing the property of expressing a truth to a sentence.
43. Bolzano, *WL I*, 206. This is another point where Bolzano disagrees with Frege: In Frege (1923) 44 the double negation of a proposition *x* is said to be identical with *x*. In Frege (1919a) 157 he still regarded them as different. Cp. the commentary on these passages in Künne (2010a).
44. I would have liked to add: 'and of the proposition that there is at least one prime number', for that seems right, too. But alas! it is unclear whether Bolzano would accept it, for it clashes with his claim that the proposition that *there is* at least one object is a truth that has no ground (Bolzano, *WL II*, 375).
45. As Roski 150 f explains, Bolzano argues in *WL II*, 399 for a thesis that (combined with his hypothesis that every truth can be expressed by a sentence of the form '(Every) A has (property) b') entails this contention.

46. Bolzano, *WL I*, 144–153, 169–200. He used to begin his courses in the philosophy of religion by going through some of the arguments against scepticism that he was to present on a larger canvas (and with greater precision) in *WL*: see Bolzano (1834a) I, 32–41.
47. See (1813) 135–138, §2; *WL III*, 453–454, *WL IV*, 449. I have added the final disjunct. I can claim his consent for this addition, for some statements in *WL* imply that the proposition that not not *p* is more complex than the proposition that not *p*. Cp. *WL I*, 355.
48. Cp. *WL II*, 362–364. Surprisingly Bolzano assigns to ASYMMETRY the status of a theorem. In his Reductio-ad-absurdum argument for ASYMMETRY the *absurdum* is the statement that the ascription of truth to a truth *x* is a ground of *x* (*WL II*, 362, *sub* 2). Tatzel (2002) takes this as evidence for his claim (p. 11) that for Bolzano the negation of this proposition is an axiom of his theory of grounding (op. cit. 14 and notes 50 f). The evidence is fairly indirect, and one wonders why Bolzano did not put this alleged axiom side by side with T-GROUND, as Tatzel does. Be that as it may, the theory becomes more transparent when one takes ASYMMETRY to be an axiom. Then the non-reversibility of 'Every truth *x* is a ground of the ascription of truth to *x*' follows immediately. The same stance is taken in Roski (2017) 81 f, 94.
49. *Met.* Θ 10, 1051 b 6–8, *Cat.* 12, 14 b 18–19.
50. Bolzano regards this principle as an axiom.
51. Consider 'He must have expected it to rain, because he took his large umbrella along.' When 'because' is used evidentially, one often encounters an epistemic 'must' in the main clause, and one tends to articulate dissent along the following lines: 'But that doesn't mean that he expected rain. He likes to use that umbrella as a walking stick.' The explanatory 'because' would be in play if the speaker had rather said, 'He took his large umbrella along because he expected it to rain.' Schnieder (2015a) contains on pp. 148–150 a helpful discussion of this distinction with several references to the pertinent linguistic literature.
52. A purely syntactical conception of substitution instances also lets us down when the connectives are truth-functional: in Strawson (1957) the test case contains two tokens of a lexically ambiguous sentence, in Künne (2008a) 300–303 the second test case contains two tokens of a grammatically ambiguous sentence. Schnieder noticed that both moves can be applied to putative counter-examples to IRREFLEXIVITY©. Since counter-examples to IRREFLEXIVITY© are *eo ipso* counter-examples to ASYMMETRY©, he puts them under close scrutiny in the course of defending ASYMMETRY© in his (2015a) 144–147.

53. I have borrowed this delightful example from Schnieder, loc. cit. Obviously, in my clumsy reformulation of Tucholsky's witticism the connective is no longer flanked by tokens of the same sentence.
54. In Künne (2003) 154–155 my example was 'He is your first cousin because he is a child of a sibling of one of your parents.' On the last page of (1997) and again on p. 46 of (2003) I turned T-GROUND without any further ado against Frege's Identity Thesis. 'Because' statements like the 'first cousin' example discussed in the very same book show that this attack was too hasty.
55. In Künne (2003) 369–371, I gave Mates-inspired reasons for denying the second assumption. I no longer find them convincing.
56. Frege (1891) 11. It should not go unmentioned that Frege makes this point in a context that turned out to be very embarrassing. The two sentences that are said to express the same thought instantiate the two sides of what was to become the fatal Axiom V of *Grundgesetze*. Restricting what Axiom V says about all functions to the special case of Fregean concepts, the axiom states: 'For all concepts *c*, *d*: everything that falls under *c* also falls under *d*, and vice versa, if, and only if, the extension of *c* = the extension of *d*'.
57. Sometimes a difference in logical form between two compound sentences is, as Frege put it in his last publication, 'an inevitable consequence of the difference between spatio-temporal phenomena and the world of thoughts' [Frege (1923) 39, cp. 40–41; Künne (2010a) 608]. When we express a compound of two thoughts we first utter one sentence-token and then another one, or we inscribe one sentence-token here and another one there. The difference between corresponding instances of '*p* and (or) *q*' and '*q* and (or) *p*', Frege pointed out, is only due to our expressive predicament. In such cases, two sentences though differing in logical form do not articulate the thought they both express in logically different ways.
58. For a study in depth of this problem see Schnieder (2010). He does not present it as a difficulty for Bolzano's theory. What follows in this paragraph is essentially the treatment of 'because' sentences like (j) that he favours: op. cit. 337–340.
59. Note that the main clause of (k) is not meta*linguistic* like that of '5 is called "prime" because etc.' but meta-*conceptual*. Unlike its metalinguistic counterpart it has the same modal and epistemic properties as the main clause of (j): what they express is true in all possible worlds, and it is knowable a priori. Schnieder (2010) 332–335 critically discusses the metalinguistic variant of the option I take to be a serious competitor of the thesis that (at least sometimes) grounds are not propositions but rather propositions as articulated in a certain way.

60. Bolzano, *WL II*, 85. (In the Oxford translation the second conjunct in the last sentence remains untranslated.) Cp. also *WL III*, 79.
61. I am very grateful to Edgar Morscher for his clarifying comments on a precursor of Sect. 6. I am particularly indebted to Benjamin Schnieder and Stefan Roski: their searching questions and constructive suggestions greatly helped me to improve the penultimate version of Sects. 3–6.
62. Cp. Bolzano (1839) 66. Bolzano (E&P) contains on pp. 199–230 a translation of the sections on morality in Bolzano (1834a) I, 227–266. (At one point in the passage quoted above the translators commit a serious mistake: under [2] they render the 'namely' rider as 'not those that are determined by God's behaviour' [(E&P) 214].)
63. For another application of this falsification strategy see Bolzano (1834a) I, 244. Unbeknownst to Bolzano, the archetype of all criticisms of divine-command views of moral goodness is to be found in Plato's *Euthyphro* 9d-11b. The connective '*ὅτι* (because)' plays in Plato's argument a role that is analogous to that of '*weil*' in Bolzano's argument. Thanks to Plato there is at least literature that is indirectly about this aspect of Bolzano's theory of grounding: the analysis of the *Euthyphro* argument in Sharvy (1972) and in Schnieder (2015b).

Appendix

While it has very often been observed that Bolzano's conception of grounding plays a key role in his work in mathematics and in the philosophy of mathematics, its role in his philosophy of religion and morality is seldom if ever taken into account. Unlike many other theists Bolzano is convinced that our moral obligations do not at all depend on God's will. Here is his argument:

> The proposition that everything that God … commands … is morally good, is not the highest moral law, [A] because not <u>all</u> practical truths can be derived from it, namely not those that determine God's behaviour, and also [B] because those that can be derived from it do not objectively flow from it as consequences from their ground. For it is not because God … commands something that we ought to will it, but rather conversely, God … commands it because we ought to will it. [*Der Satz an sich,*] *daß Alles, was Gott … gebietet … sittlich gut sey,… ist … nicht das oberste Sittengesetz,* [A] *weil sich nicht <u>alle</u> praktischen Wahrheiten, nämlich nicht diejenigen, die das Verhalten Gottes selbst bestimmen, aus ihm herleiten lassen, und* [B] *weil auch diejenigen, die sich aus ihm herleiten lassen, aus ihm nicht objectiv, nicht wie die Folge aus ihrem Grunde, fließen. Denn nicht darum, weil Gott Dieß*

> *oder Jenes … gebietet, soll es von uns gewollt werden; sondern umgekehrt, weil es von uns gewollt werden soll, … gebietet es uns Gott.* (1834a) I, 247; '[…]' inserted[62]

[A] is supposed to show that the predicates 'is commanded by God' and 'is morally good' are not even coextensive. Bolzano assumes that God's essence is such that He cannot be the addressee of commands, and yet His actions are always morally good. The interesting point is [B]. Suppose that whatever is commanded by God is morally good, and we have somehow found out that God commands us to try to help somebody who is in need of help. Then the conclusion follows that it is morally good to try to help that person. But even if the premises of this deductively impeccable little argument were true it would not present grounds for the truth that it is morally good to try to help that person. For it is not the case that such an action is morally good because it is commanded by God, but rather conversely, it is commanded by God because it is morally good anyway. Quite generally, take any general term 'F' that is declared to be extensionally equivalent with 'morally good', the observation 'Nothing is morally good because it is F, but rather conversely, something is F because it is morally good' falsifies the claim 'The statement that something is morally good if, and only if, it is F is the supreme moral law' even if that statement is true.[63]

References

Aquinas/Tommaso d'Aquino. (c. 1271) 1950. *In Libros Duodecim Metaphysicorum Aristotelis Expositio*, ed. M.R. Cathala and R.M. Spiazzi. Torino: Marietti.

Aristotle. 1949. *Categoriae et Liber de Interpretatione*, ed. L. Minio-Paluello. Oxford: Clarendon Press.

Aristotle. 1953. *Metaphysics*, ed. Ross. Oxford: Oxford University Press.

Beaney, Michael, and Erich H. Reck (eds.). 2005. *Gottlob Frege—Critical Assessments*, vol. I. London: Routledge.

Bolzano, Bernard. 1813. Neue Theorie der Parallelen. In (BGA) 2A.5, 135–138.

Bolzano, Bernard. 1834a. *Lehrbuch der Religionswissenschaft*, 4 vols. In (BGA) 1.6/1–8/4.

Bolzano, Bernard. 1834b. Letter to Franz Exner. In (BGA) 3.4/1.

Bolzano, Bernard. 1834c. Verbesserungen und Zusätze zur *Logik*. In (BGA) 2A.12/2.

Bolzano, Bernard. 1837. *Wissenschaftslehre*, 4 vols. Repr. Aalen 1981. In (BGA) 1.11/1-14/3. Trans. as *Theory of Science*, ed. and trans. R. George and P. Rusnock. Oxford: Oxford University Press, 2014.

Bolzano, Bernard. 1839. *Dr. Bolzano und seine Gegner*. In (BGA) 1.16/1.

Bolzano, Bernard. 1851. *Paradoxien des Unendlichen*, Hamburg: Meiner, 2012. Trans. as *Paradoxes of the Infinite*, ed. and trans. D.A. Steele. London: Routledge and Kegan Paul, 1950.

Bolzano, Bernard. 1969 ff. (BGA) *Bolzano Gesamtausgabe* [Opera omnia], ca. 126 vols., ed. Jan Berg et al. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Bolzano, Bernard. 2007. (E&P) *Selected Writings on Ethics and Politics*, ed. and trans. P. Rusnock and R. George. Amsterdam: Rodopi.

Buridanus/Jean Buridan. 1977 (c. 1350). *Sophismata*, ed. T.K. Scott. Stuttgart: Frommann-Holzboog.

Chisholm, Roderick M. 1960. *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice Hall.

Dummett, Michael. 1978 (1959). Truth, as repr. in his *Truth and Other Enigmas*, 1–24. London: Duckworth.

Dummett, Michael. 1981. *The Interpretation of Frege's Philosophy*. London: Duckworth.

Dummett, Michael. 1991. *Frege—Philosophy of Mathematics*. London: Duckworth.

Dummett, Michael. 1997. Comments on Wolfgang Künne's Paper. *Grazer Philos. Studien* 53: 241–248. Repr. in Beaney and Heck (2005).

Dummett, Michael. 2007. Reply to Wolfgang Künne. In *The Philosophy of Michael Dummett*, ed. R.E. Auxier and L.E. Hahn, 345–350. Chicago: Open Court.

Frege, Gottlob. c. 1883. *Kernsätze zur Logik* [Key Sentences on Logic]. In (NS//PW).

Frege, Gottlob. 1891. *Funktion und Begriff* [Function and Concept]. In (KS//CP).

Frege, Gottlob. 1892. *Über Sinn und Bedeutung* [On Sense and Reference]. In (KS//CP).

Frege, Gottlob. 1893. *Grundgesetze der Arithmetik*, Vol. I; (1903): Vol. II. Repr. Hildesheim: Olms, 1998, trans. as *Basic Laws of Arithmetic*, ed. and trans. Ph. Ebert and M. Rossberg. Oxford: Oxford University Press, 2013.

Frege, Gottlob. 1897. *Logik* [Logic]. In (NS//PW).

Frege, Gottlob. 1913. *Vorlesung über Begriffsschrift* (Carnap's Jena Lecture Notes, ed. G. Gabriel). *History and Philosophy of Logic* 17 (1996): 1–48.

Frege, Gottlob. 1914. *Logik in der Mathematik* [Logic in Mathematics]. In (NS//PW).

Frege, Gottlob. 1915. *Meine grundlegenden logischen Einsichten* [My Basic Logical Insights]. In (NS//PW).

Frege, Gottlob. 1918. *Der Gedanke* [The Thought]. In (KS//CP).
Frege, Gottlob. 1919a. *Die Verneinung* [Negation]. In (KS//CP).
Frege, Gottlob. 1919b. *Brief an Paul F. Linke*. In (WB//PMC).
Frege, Gottlob. 1923. *Gedankengefüge* [Compound Thoughts]. In (KS//CP).
Frege, Gottlob. (NS//PW) *Nachgelassene Schriften*, ed. H. Hermes et al., Hamburg: Meiner, 1969//*Posthumous Writings*, trans. P. Long and R. White. Oxford: Blackwell, 1979.
Frege, Gottlob. (WB//PMC) *Wissenschaftlicher Briefwechsel*, ed. G. Gabriel et al., Hamburg 1976//*Philosophical and Mathematical Correspondence*, ed. B. McGuinness. Oxford: Blackwell, 1980.
Frege, Gottlob. (KS//CP) *Kleine Schriften*, ed. I. Angelelli, Darmstadt ^{2}1990//*Collected Papers*, ed. B. McGuinness. Oxford: Blackwell, 1984.
King, Jeffrey C. 2007. *The Nature and Structure of Contents*. Oxford: Oxford University Press.
Künne, Wolfgang. 1997. Propositions in Bolzano and Frege. In *Grazer Philos. Studien* 53. Repr. in Beaney and Reck (2005) and in Künne (2008b), 157–196.
Künne, Wolfgang. 2001. Constituents of Concepts: Bolzano vs. Frege. Repr. in (2008b) 211–232.
Künne, Wolfgang. 2003. *Conceptions of Truth*. Oxford: Clarendon Press.
Künne, Wolfgang. 2007. Two Principles Concerning Truth. In *The Philosophy of Michael Dummett*, ed. R.E. Auxier and L.E. Hahn, 315–344. Chicago: Open Court.
Künne, Wolfgang. 2008a. Analyticity and Logical Truth. From Bolzano to Quine. Repr. in (2008b), 233–304. Sankt Augustin: Academia Verlag.
Künne, Wolfgang. 2008b. *Versuche über Bolzano/Essays on Bolzano*. Sankt Augustin: Academia Verlag.
Künne, Wolfgang. 2008c. The Modest, or Quantificational, Account of Truth. In *Truth= Studia Philosophica Estonica* 1, ed. D. Cohnitz, 122–168.
Künne, Wolfgang. 2010a. *Die Philosophische Logik Gottlob Freges*. Frankfurt: Vittorio Klostermann.
Künne, Wolfgang. 2010b. Reply to Paul Boghossian. *Dialectica* 64: 585–597.
Künne, Wolfgang. 2013. Truth Without Truths? 'Propositional Attitudes' Without Propositions? Meaning Without Meanings? In *Studies in the History and Philosophy of Polish Logic*, ed. K. Kijania-Placek et al., 160–204. London: Palgrave Macmillan.
Künne, Wolfgang. 2015a. Frege on That-Clauses. In *Dummett on Analytical Philosophy*, ed. B. Weiss, 135–173. London: Palgrave Macmillan.
Künne, Wolfgang. 2015b. On Having a Property. Corrigenda in Bolzano's *Wissenschaftslehre*. *Grazer Philos. Studien* 91: 365–408.
Morscher, Edgar. 2014. Bernard Bolzano. In *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), ed. Edward N. Zalta. Stanford: Stanford University Press. https://plato.stanford.edu/archives/fall2014/entries/bolzano/.

Quine, Willard Van Orman. 1970. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice Hall.
Roski, Stefan. 2017. *Bolzano's Conception of Grounding*. Frankfurt: Vittorio Klostermann.
Schnieder, Benjamin. 2010. A Puzzle About "Because". *Logique et Analyse* 211: 317–343.
Schnieder, Benjamin. 2015a. The Asymmetry of "Because". *Grazer Philos. Studien* 91: 131–164.
Schnieder, Benjamin. 2015b. Das Meisterargument in Platons Euthyphron. *Philosophiegeschichte und Logische Analyse* 18: 227–254.
Sharvy, Richard. 1972. Euthyphro 9d–11b: Analysis and Definition in Plato and Others. *Noûs* 6: 119–137.
Strawson, Peter F. 1950. On Referring, as repr. in his 1971, 1–27.
Strawson, Peter F. 1957. Propositions, Concepts, and Logical Truths, as repr. in his 1971, 81–100.
Strawson, Peter F. 1971. *Logico-Linguistic Papers*, 81–100. London: Methuen.
Tatzel, Armin. 2002. Bolzanos's Theory of Ground and Consequence. *Notre Dame Journal of Formal Logic* 43: 1–26.

The Names of the True

Paolo Leonardi

On the seventh page of "Über Sinn und Bedeutung" ("On Sense and Meaning"), Gottlob Frege turns to discuss the meaning of a sentence. About two pages later, he writes:

> We have seen that the meaning of a sentence may always be sought, whenever the meaning of its components is involved; and that this is the case when and only when we are inquiring after the truth-value.
>
> We are therefore driven into accepting the *truth-value* of a sentence as constituting its meaning. By the truth-value of a sentence I understand the circumstance that it is true or false. There are no further truth-values. For brevity, I call the one the True, the other the False. Every assertoric sentence concerned with the meaning of its words is therefore to be regarded as a proper name, and its meaning, if it has one, is either the True or the False. These two objects are recognized, if only implicitly, by everybody who judges something to be true – and so even by a sceptic. (1892: 33–34 [1997: 157–58])[1]

I am grateful to Sergio Bernini, Wolfgang Carl, Pierdaniele Giaretta and Marco Santambrogio, and to the Padova seminar in analytic philosophy, for their careful reading and feedback.

P. Leonardi (✉)
Department of Philosophy and Communication Studies,
University of Bologna, Bologna, Italy
e-mail: paolo.leonardi@unibo.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_4

Frege calls an expression that stands for an object, *proper name* (hereafter, name); the object it stands for is, its *meaning*; and what determines its meaning is, its *sense*. A name expresses its sense.[2] The opening line of the quoted passage specifies when one can seek for the meaning of an assertoric sentence (hereafter, sentence). Frege then states that the sentence's truth-value is its meaning—its being true or false, the True and the False, which he qualifies as logical objects. Hence, the meaning of a sentence, "if it exists, is either the True or the False," an object, while the sentence is itself a name.

The thesis that a sentence is a name, surprising as it is, requires some support. The idea is also briefly introduced in Frege 1891, published one year earlier, but written after "On Sense and Meaning." Some have claimed that Frege does not supply enough backing for his argument. That a sentence be a name, it has been added, does not relate to our intuitions, and there is no hint that, nonetheless, we ought not resist it.[3] Scholars have connected the thesis with Frege's 1893 ideography, where it plays no technical role.[4] Michael Dummett, perhaps the best known Fregean of the last fifty years, was very critical of the thesis. For Dummett, in undertaking this move, Frege renounced the context principle—"[…] never to ask for the meaning of a word in isolation, but only in the context of a proposition"[5]—the hinge on which he had structured semantics in his work of 1884. The picture, I shall argue, was drawn to fit the formal project, but it respects our pre-theoretical intuitions and does not undermine the sentence's semantic role. Frege's paradigm of the proper name (a phrase he uses almost as 'singular term' was traditionally used) is the definite description (hereafter, description). A description wears its sense on its sleeves and, if it describes an existent object, that object is its meaning. Already in the *Begriffsschrift* (*Conceptual notation*) of 1879,[6] Frege connects sentences and definite descriptions when he nominalizes the sentence 'Archimedes perished at the capture of Syracuse' as 'The violent death of Archimedes at the capture of Syracuse'. What this description means may be what that sentence means, but what kind of object does 'The violent death of Archimedes at the capture of Syracuse' mean? And if the description and the sentence do not mean the same object, what kind of object does the sentence mean? I shall suggest an answer to both questions, a suggestion that is compatible with the constraint that all true sentences mean the same. If "in the referent of the sentence all that is specific is obliterated" ([1892: 35] 1948: 217), the object that I take them to

mean is compatible with Frege's view. Later, I shall tackle the meaning of false sentences.

Eva aligned with Dummett, adding that Frege 1892 requires many different naming relations for classifying sentences as names. Eva was a Fregean, whereas I am not. My approach here is not to offer an interpretation of Frege, but rather an external understanding of his work.[7] If Eva could have read the paper, she would have corrected me with her serious irony, and I would have understood better.[8]

1 Articulating an Idea

Bertrand Russell is among the first to have resisted viewing sentences as names. On February 20, 1903, he writes in a letter to Frege:

> I have read your essay on sense and reference, but I am still in doubt about your theory of truth-values, if only because it appears paradoxical to me. I believe that a judgement, or even a thought, is something so entirely peculiar that the theory of proper names has no application to it. (PMC: 155–56)

Russell's outlook on the issue was different—he never conflated proper names and descriptions, and in the *Principles of Mathematics*, published the same year, 1903, he ascribed names the role of indicating a particular. In the letter written in response to Russell's, three months later, on May 21, 1903, Frege repeats part of what he had written eleven years earlier, with no additions. Specifically, he insists that the sentence meaning isn't the thought it expresses and that if the meaning of its parts is relevant to a sentence as their sense is, we have to look elsewhere for the meaning of the sentence. What can there "be other than its truth-value[?]" Why not occurring states of affairs? Or, as Russell himself might have objected, why not facts? To state that "a fact is a thought that is true," is to exclude facts by definition.[9] Frege is *articulating*, not arguing for, an alternative picture. His outline can be presented as a frame for both his formal program and our pre-theoretical intuitions.

"Let us *assume*, for the time being," begins Frege, "that the sentence has a meaning!" (1892: 32 [1948: 214–15], my italics.) Then, he examines whether the thought that a sentence expresses is "its sense or its meaning." It is not the sentence meaning because by substituting a

word with another one with the "same meaning, but a different sense," the thought changes but the sentence meaning does not. Hence, the thought is the sense of a sentence, not its meaning. In a few pages, the idea that a sentence has a meaning moves from an assumption towards a matter of fact, and its meaning comes to take the specific form of a truth-value. We "recognize and expect" there to be a meaning of the sentence, and we look for it because "we are concerned with its truth-value." "[S]triving for truth" we "advance from the sense to the meaning" (1892: 32 [1948: 215–16]). We can dwell on a thought without worrying what it is about, for instance, when reading fiction, going to the movies, watching an actress perform, or fantasizing by ourselves. But when we want to know what is the case, the matter changes, and we take greater care. As such, hopefully, we shall come to know what it is we look for—whether the sentence is true or false. To pursue sentence meaning is to pursue truth.

With a further push, the idea comes closer to being true by definition. We seek the meaning of a sentence, Frege adds, when "the meanings of its components are involved" and "this is the case when and only when we are inquiring after the truth value" (1892: 33–34 [1948: 216]). Our wanting to know the meanings of a sentence's components depends on our wanting to know the meaning of the sentence, its truth-value, whether it is true or false.[10] In being instrumental—i.e., looking for the meaning of a sentence so as to come to know its truth-value—we come to pursue one and the same thing.

Lastly, however, Frege seems to offer an argument derived from Leibniz's principle of substitution *salva veritate*.

> If our supposition that the meaning of a sentence is its truth value is correct, the latter must remain unchanged when a part of the sentence is replaced by an expression having the same meaning. And this is in fact the case. Leibniz explains: "*Eadem sunt, quae sibi mutuo substitui possunt, salva veritate.*" What else but the truth value could be found, that belongs quite generally to every sentence concerned with the meanings of its components and remains unchanged by substitutions of the kind in question? (1892: 35 [1948: 217])

The application of Leibniz's principle is topsy-turvy. On Leibniz's principle, for instance, 'The author of *De officiis*' can be substituted with 'The author of *De amicitia*' in any (non-oblique context) sentence

without changing the truth-value of the sentence, because they mean the same man, Cicero, and hence *eadem sunt*. But instead, from the principle Frege derives that the truth-value of a sentence is its meaning, since that value does not change if a term in the sentence is substituted with another one that has the same meaning. Assuming this does not imply that one can go in the reverse direction, i.e., it does not follow that a sentence truth-value is its meaning because substituting one part in a sentence with another with the same meaning, the sentence truth-value does not change. The sentence meaning could be something else. Here, again, is an instance of that 'something else': (ii) The thought is classified as the sense, not the meaning, of a sentence because by substituting a word in a sentence with another with a different sense but the same meaning, it is claimed that the thought the sentence contains varies, while its meaning remains the same. Could one not instead maintain that the meaning of the sentence changes too? For instance, one would contend that 'The author of *De officiis* was a Roman senator' and 'The author of *De amicitia* was a Roman senator' have a different meaning because an element of the first is the *De officiis* and an element of the second is the *De amicitia*. Russell could have argued thus, both when he took a true sentence to correspond to a fact and when he later adopted the multiple relation theory of judgment.[11] Frege's Leibnizian reflection has a rhetorical flavor: The principle says nothing about the meaning of a sentence as depending on the meaning of its parts.

I shall now touch on each of the three points mentioned at the end of this first section, starting with the second one—there are proper intuitions and ways of looking at the matter that make it more palatable. I shall then move on to the role a sentence plays in Frege's formal system.[12] I shall end with a short argument countering the supposed displacement of the sentence, if a proper name, from the center of semantics.

2 A Fregean Proper Name

According to Frege, any expression, simple or composite, which means an object, provided such a designation exists,[13] is a proper name. Consequently, Frege applies the label *proper name* to pick out both actual proper names and definite descriptions. As he writes:

> The designation of a single object can […] consist of several words or other signs. For brevity, let every such designation be called a proper name. (1892: 27; 1948: 210)

We name cities, rivers, mountains, countries, wars, chemical elements, days, happenings—'the Armistice,' 'the capture of Syracuse'—ceremonies—'Elizabeth I's coronation,' etc. Nonetheless, Frege's idea, and ours, of what an object is seems clear and reasonably coherent. A Fregean proper name can mean any such single element,[14] whereas, traditionally, a singular term *designates*—designation being a more imprecise relation than *Fregean meaning*. An implicit argument for bringing them together is that both an actual proper name and a definite description can occur at either side of the identity sign, as in '*Venus* is *the morning star*.'[15] Between an actual proper name and a definite description, though, there are two important differences. First, a single object may be given many actual proper names. Frege himself uses 'Venus,' 'Phosphorus,' 'Hesperus' to name the same celestial body. Another example that has entered the philosophical debate is Mount Everest (whose name comes from the surname of a Welsh colonel and geographer), but whose Tibetan (and Chinese) name is 'Chomolungma,' and its Nepali one is 'Sagarmāthā'—again, capitalized description but unknown to most westerners. Some Europeans understood that the mountain was also called 'Gaurishankar'—not surprised that the mountain possibly had a fourth actual proper name.[16] But, in any language, indefinitely many definite descriptions fit a single object. There is no problem in speaking of Mary's house as the second villa on the left after the traffic lights going towards Siena, the house where Bob used to live when young, the best investment in town, the place I spent New Year's Eve 2015, Adelaide's nightmare. The house has many features, and we can designate it by availing ourselves of any (combination) of them. It has spatial relations with any other object in the universe, whose center of mass and direction at time t can be located. Moreover, we can designate the number 8 also by '4×2', '2^3' or '$5 + 3$', and so on. Or, to take a different case, a form of *crescendo*: the buttering of a piece of toast, the slow buttering of a piece of toast, the slow and deliberate buttering of a piece of toast, the slow and deliberate buttering of a piece of toast in the bathroom…[17] Second, whereas "in the case of an actual proper name such as 'Aristotle' opinions as to the sense may differ,"[18] this seems not to be the case with definite descriptions.[19] The problem with the *sense* of an actual proper

name has been an issue ever since. Frege patches the problem in the following way, asserting that:

> [s]o long as the referent remains the same, such variations of sense may be tolerated, although they are to be avoided in the theoretical structure of a demonstrative science and ought not to occur in a complete language. (1892: 27 fn 2 [1948: 210, fn 2])[20]

Nothing so far suggests classifying a sentence too as a proper name. Can we, in part at least, build up a case? Any Fregean proper name can occur on the left or the right of the identity sign, can a sentence occur in either place? At first glance, the answer would seem negative. 'York is an English city is identical to New York is an American city' looks hard to make sense of. If the claim is that 'York is an English city' has the same truth-value as 'New York is an American city,' however, it does make sense. Besides, already in the *Begriffsschrift*, with the sentence 'Archimedes died' and the description 'Archimedes' death,' we can see Frege connecting sentences and descriptions. If the second has a meaning, the first cannot lack one—they can mean the same or two related objects (after all, Frege claims that *the concept horse is not a concept*), taking into account that we name any object. The relation between sentences and descriptions, however, goes beyond the nominalization of sentences. In English, we can rewrite *the daughter of Mary* as *The individual such that* (*she is a daughter of Mary*), and in semi-formal English we can, in general, rewrite *the F* as ιx (*x* is *F*). That does not eradicate the differences between 'Archimedes died' and 'Archimedes' death'[21]: the first one predicates something of Archimedes, while the second uses that predication to pick out the event of his death. Frege could have held that the description presupposes the sentence—in the description, the focus is not on the predication, which this time is put to work to point out an event. As there are indefinitely many descriptions of the same object, what a sentence is about can be what is about an indefinite number of sentences—think of 'She was buttering a piece of toast,' 'She was slowly buttering a piece of toast,' etc. As the sense of a description seems transparent, so too does that of a sentence.

A description may mean any object, from a speck of dust to the universe; however, a sentence, if it means at all, cannot mean most objects, from the speck of dust or of blue up. A sentence means not a speck of dust, but possibly the speck of dust being blue. A naïve view would have

it that sentences describe facts, but Frege himself takes them to be true propositions—"a fact is a thought that is true," he writes.[22] Other times, facts are accounted for as states of affairs that obtain. Sentences can be said to describe states of affairs rather than facts because they can be about something that does not obtain—and, in this case, they are false. Combining the possibility of describing the same object in indefinitely many ways and Frege speaking of two objects only as what a sentence means, we could maintain that all true sentences mean the substance, the world, what is—the amalgam of what is. All false sentences misdescribe the substance, the world,[23] what is. I would add, a false sentence means nothing, but we manufacture a meaning for it, the False. Nature or god is Spinoza's only substance—the world or what is are secular alternatives.

If this could be elaborated into a coherent picture that could make Frege's idea somehow intuitive, yet it isn't Frege, who has no nature, world or what is as sentence meaning.

3 The Fregean Sentence as a Proper Name

A sentence, according to Frege, means its truth-value, i.e., either the True or the False. The True and the False are, he adds, "logical objects." Many scholars have linked his view of sentences as proper names to Frege's formal project in *The Basic Laws of Arithmetic*. Frege's choice may be seen as pragmatic,[24] or simply as a convenient way of speaking.[25] But it is more than that. It focuses on the *point* of logic—looking for the true consequences of true premises. This is a relation between true sentences, for which the distinguishing of sentences that are true from those that are not is fundamental. In choosing the True and the False as the objects a sentence means, and calling them *logical* objects, Frege upholds the basic understanding of what logic is, insofar as a formal study is indifferent to any specific material, opting for an *epoché*, a suspension of judgment, concerning any informal detail.[26] If suitable, logical reasoning can be applied to a real case, substituting English singular terms for formal language individual variables and English predicates for formal language predicate constants—a step which does not pertain to logic.

In sticking to two truth-values, Frege displays a philosophical preference—his logic is a truth-based one, and he is interested only in a sentence being true or not. Hence, his positing only two values. If, contrary to Frege, you want to distinguish between three values or more,

reasoning as he does, you would acknowledge as many logical objects as sentence meanings. Indeed, one might acknowledge, for instance, four values—say, {Ø, {T}, {F}, {T,F}}. A reading of the four values is: *no information concerning this state of affairs*; *information saying that the state of affairs fails*; *information saying that the state of affairs obtains*; *conflicting information saying that the state of affairs obtains as well as fails*.[27]

Beyond the formal project, it is debatable whether the True and the False are on the same footing. In the formal project, too, they are not exactly on a par. In Frege's ideography, the True is the basic option: any sentence claims its truth, and the only sign of judgment he uses is that of assertion, the vertical. The False is a misjudgment. The assertion *not p* is the only way to represent that *p* is false—something which intuitively sounds like: truly, *p* is false.[28]

Frege, I think, believes his ideography to be largely compatible with our standard talk exchanges. In the pages we are analyzing, this view appears in at least two passages. The first is at the close of the first passage I quoted: "These two objects [the True and the False] are recognized, if only implicitly, by everybody who judges something to be true." The second one, which I also quoted above, occurs one page later. There, Frege asserts that the True or the False is the meaning of a sentence if in it "all that is specific is obliterated" (1892: 35 [1948: 217]). But he has no concern for, or does not want to enter into, a philosophical and controversial ontological debate on this theme. It seems natural to match a sentence to a fact or a constituent of one if a fact is an event under description. Frege, as we have seen, excludes this match. Be that as it may, the identity of a fact at a close distance turns out to be troublesome. There is an event, which is this or that fact depending on how it is described. Which fact am I part of now? I am at home, I am in Florence, I am in Italy, … I am writing, I am writing on my computer, I am writing on my computer using Word, I am writing a paper on my computer using Word … Are these distinct facts? What is distinct in these facts? The two matryoshka series, the two series themselves, a fact of the first series and one of the second? Am I not at home, writing? The event on the back of the fact hardly fares any better. Usually, we hold a fact not to be an object—"The world is the totality of facts, not of things," states Wittgenstein almost 30 years after the publication of "On Sense and Meaning." Anyway, Frege is strict in separating objects from other entities, and facts were never on his list of entities either. He devises the objects to which a

sentential name refers along a different path without making any general ontological claim analogous to Wittgenstein's. Frege builds up the ontology he needs, the True and the False, two logical objects, two abstract entities.

4 *The False More at Large*

My claim is that Frege's technical view of a sentence meaning is compatible with my wider and more vague picture. If so, his view relates to our intuitions. Above, however, I explicitly considered only true sentences—which, I maintain, can be taken to mean what is or nature. In the previous section, I remarked that, in Frege, the True and the False are asymmetric: the first is the central case on which the second depends. How do we look at *sentences meaning the False* in the more vague picture?

In 1914, Russell lectured at Harvard, "I argued that there were negative facts," he reports, "and it nearly produced a riot: the class would not hear of there being negative facts at all" (1919: 42).[29] Frege's maintaining that some sentences mean the False, and that the False is a logical object along with the True, did not elicit any such turmoil. This is mainly because the False is not an ontological posit, nor is backed by one. The False is a coin of account, since a false thought or the false statement that gives it expression are managed by the truth of their negation, or by their negation denoting the True. The True, I think, is a coin of account too, put to the task of skipping over the details of what is. However, contrary to the False, the True is backed by an ontological posit. If beyond the True there is what is, or nature, beyond the False there is nothing. Hence, notwithstanding the False, Frege's approach is radically different from that of Russell.[30]

The asymmetry in Frege's technical picture is paralleled in the fuzzier one. '*p*,' if true, matches *p*, and '*p*,' if false, matches nothing. In a simple case, 'not *p*' is false because of *p*. Other cases are more complicated because there are more alternatives, as happens with colors. If it is not red, it may be blue, green, yellow, etc.; if it is blue, it is not red. Schematically, then, if it is *not p*, it may be *q*, *r*, *s*, *t*, … Since the basic colors are three (red, blue, green), and the primary ones are six (the three basic colors plus yellow, black, and white),[31] one has to check for the alternatives.[32] Some cases are complicated because we do not

even know exactly how many alternatives there are. 'Desdemona loves Cassius,' if false, matches a range of meanings: her being indifferent to him (including the chance that she does not even know him); her hating him; her having some sympathy for him but no more; her engaging in some flirting simply because she has a seductive temperament, etc. In any case, the idea is that something is false because some alternative is true.

In general, the case of a false sentence can be looked at in the following way: a sentence is false because it matches nothing—just as a singular term is vacuous because it matches nothing. This demonstrates a parallelism between singular terms and sentences, and between reference and truth.[33] It is a parallelism with a difference: in the false sentence case, with a caveat, there is something that mismatches the false sentence, while in the case of the vacuous term, there is nothing that it can even mismatch with. In other words, if a singular term is vacuous there is nothing there; if a sentence is false, and no vacuous term occurs in it, something else is true instead. Negation is a tool to represent what is not, a representation that could itself be false. The False—as possible but not actual circumstances, and as impossible ones too—is only a representation; it is not part of what is. It may be convenient, though, to take it as an abstract object, or more precisely as a logical object: there are two cases with a representation—it either fits what is, or it does not. In the second case, one could be inclined to anticipate the negative frame, speaking of a *misrepresentation*.

The parallelism between vacuous terms and false sentences shows the effect that a vacuous term has on the meaning of the sentence in which it occurs—the caveat I hinted at above. A requirement for judging a sentence true or false is that each of its components has a meaning. A sentence in which a vacuous name, or one satisfied by more than one individual or one object, occurs, according to Frege, either has no truth-value—a representation without a represented—or it speaks of the set of elements that the vacuous term names. The second is the technical solution Frege adopts in *Basic Laws*. 'The author of the three volumes of *Principia Mathematica* is an English native speaker' gets as meaning for its argument the set whose members are Bertrand Russell and Alfred N. Whitehead. 'The author of the three volumes of *Principia Philosophica* is an English native speaker' gets the null set. In either case, the sentence turns out false.[34]

5 The Semantic Role of the Sentence

Following Dummett, Eva complains that if the sentence is a proper name, it loses the central role it previously played in Frege's semantics. In the *Foundations*, Frege introduces the context principle, which implies the central semantic role of the sentence: "[…] it is only in the context of a proposition that words have a meaning" (1884: 73, §62 [1960: 72, §62).[35]

In 1981, in *Assertibility and Truth*, her first book, Eva writes:

> … the different ways in which a sentence can be assessed as true show, in my opinion, that there is no such thing as a unique name-relation that obtains for all sentences. The name-bearer prototype is particularly misleading because it suggests that in all these cases we are dealing with a problem of naming and not with the particular manner in which the meaning of a sentence is related to its truth-conditions. (1981: 42–43)

Eva imagines how a speaker can recognize whether a sentence is assertible—does it have to exhibit a proof, as in mathematics, to know a test, as with an empirical statement, to know how to individuate an object, as with an existential claim? Three different manners in which to relate a sentence to its truth-conditions, three different name relations?

Returning to the issue 13 years later, she concludes that

> [t]he main objection that can be raised […] against Frege's analysis is not so much that it reifies the True and the False, but rather that it equates a sentence's role with that of a proper name. Such an identification deprives sentences of the privileged position that the context principle attributes to them. The identification is likely responsible for the least convincing developments of Fregean semantics, such as, say, the idea that the occurrence of a vacuous name in a sentence radically frustrates our intention to use it to make an assertion and presents an obstacle to articulating a coherent semantics. (1994: 120–21 [my translation])[36]

Burge hints at Dummett's and others' reaction to Frege's assimilation of sentences to proper names. Frege's critics hold the assimilation "to be an irritating peculiarity," and to them, the "claim that there are only two objects denoted by sentences and that these are Truth and Falsity" seems "to advance from the peculiar to the bizarre."[37] Burge challenges these judgments, denying that Frege had given up the context principle and the central semantic role of sentences.

> … [S]entences and terms are not everywhere interchangeable even within the formal system Frege presents in *Basic Laws*. So in a further sense, they do not have the same 'logical powers' despite the fact that they are of the same 'logical type'. Only sentences can follow the vertical judgment stroke in Frege's syntax; ordinary terms cannot. (Burge 1986: [2004: 112])

> The vertical judgment stroke is not a function sign, but is the sign of an act – judgment or assertion – an act that applies only to thoughts or sentences. (Burge 1986: [2004: 113])

Over the last 50 years and more, the notion of context has often been larger than that used by Frege, for whom the context is linguistic. I shall keep within these limits, and individual words composite expressions—i.e., sentences, definite and indefinite descriptions, complex verbal phrases are contexts. Claiming a sentence to be a proper name classifies it alongside definite descriptions. However, definite descriptions depend on sentences because they assume, imply or presuppose that someone or something satisfies their descriptive condition.[38] That assumption can be made explicit via the use of a relative or some other restrictive clause. 'The/a watermelons vendor' then would be rewritten as 'The one/One who sells watermelons,' as ιx (x is the/a watermelons vendor), or as Det x (x is the/a watermelons vendor) [where *Det* stands for *determiner*]. Besides, a sentence is syntactically distinguished by being the top syntactical category, which merges only into a sentence, but when it merges into a noun phrase as a relative clause component of it.[39] Finally, only a sentence occupying the argument place after the complex sign, |–, allows the judgment to be true.[40] Hence, even if they were proper names, sentences would not be ousted from the center of semantics.

Notes

1. In "Function and Concept", Frege writes:
   A statement contains (or at least purports to contain) a thought as its sense; and this thought is in general true or false; i.e. it has in general a truth-value, which must be regarded as the meaning of the sentence, just as (say) the number 4 is the reference of the expression "2 + 2," or London of the expression "the capital of England." (1891: 16)
2. Burge writes that "'Bedeutung' is a common word in German," (1986 [2005]: 86) as indeed 'meaning' is in English, but he does not use 'meaning' for 'Bedeutung' afterwards. See also Angelelli (1978),

Tugendhat (1970), Bell (1980). Holenstein (1983) offered a cautious evaluation of the case, but the discussion has progressed further, see Gabriel (1984). My choice of translation, i.e., 'meaning' for 'Bedeutung,' follows the lead of Long and White (1980).
'Meaning' can both be the pointing out of an expression and what is pointed out by it. Here, it has to be understood, for the components of the sentence as for the sentence itself, as what is pointed out.

3. Cf. Textor (2011): 225.
4. Frege (1893).
5. 1884: x [1960: xxii].
6. 1879: 3 [1967: 12].
7. There are many interesting analyses of Frege on this issue, besides Dummett (1973) and Picardi (1981, 1994); Burge (1984, 1986 and 2005), Ricketts (2003), Heck and May (2018), Textor (2011), Künne (2003, 2008). I particularly like Burge's, Heck and May's, and Künne's. My argument runs close to Burge (1986: 93–96). My understanding of Frege, however, is chiefly concerned with his ontology and his view of logic, rather than with his formal system and pragmatic choices.
8. Eva has been my closest colleague for 21 years, and a friend for even longer. I have come to think of her as a sister.
9. Frege (1918: 74). Davidson (1969). Already Ducasse (1940) contended that "'a fact' and 'a true proposition' mean identically the same thing" (710). See also Prior (1971: 5).
10. Frege does not consider that we may want to also fantasize about real persons and places.
11. See Lebens (2017) and Bernini (2018).
12. Burge (1984, 1986), Ricketts (2003).
13. The reservation is Frege's (1892: 32 [1948: 214]).
14. In 'The three girls were leaving early,' 'the three girls' refers to *three* single elements.
15. See Lockwood (1975) and Moro (2017).
16. In an ideal language, a single object, so Wittgenstein would later claim, would have just one actual proper name. (1921: 5.53 and its subsections).
17. Cf. Davidson (1967).
18. Frege (1892: 27), fn B (1948: 210, fn 2).
19. I doubt that the sense of a definite description may be the same for everyone if that of a name is not. There are descriptions in which an actual proper name occurs, as in 'the boy from Berlin.' Besides, do 'boy' or 'evening' have the same sense for anyone?
20. On this point, Russell dissented—names have no sense, they indicate—and, later on, Donnellan and Kripke deny a proper name a sense—names have no sense, they directly refer.

21. Mixing together parts of Frege's work that were written 40 years apart: in 1879, he would have claimed that 'Archimedes' death is a fact;' in 1918, he analyzed 'a fact' as a true proposition. Assume that a proposition is the sense of a sentence and that that-clauses refer to its sense rather than to its meaning. Then, 'It is true that Archimedes died' says the same as 'Archimedes' death is a fact.'
22. See note 10 above. I believe that a less thought-dependent understanding of Frege's claim could be offered.
23. But see Jubien (1991).
24. Burge (1986).
25. Quine (1974), Textor (2011).
26. Frege's view accounts for his insensitiveness to the slingshot problem. But so does my "intuitive" understanding of the True and the False.
27. Belnap 1977's four-valued logic, further developed into a 16-valued logic by Shramko and Wansing (2005).
28. On this topic, Burge insists at length:

> Truth is the 'objective' of judgment; the most general laws governing this objective form the subject matter of logic.
>
> […]
>
> In a sense to be explicated, truth is even more basic than falsity. The laws of logic were for Frege "nothing other than an unfolding of the content of the word 'true'". (Burge 1986: 123 [2004:])
>
> … [T]ruth is the aim of logic […]. This aim is revealed in assertion, not simply in the grammatical form of sentences. (Burge 1986: [2004: 130])

However, I would distinguish between the relevance of truth in the making up of a thought or a statement, and truth as what is asserted.

29. Around 1900, for a few years, Moore and Russell posited propositions as truth-bearers, and (hesitantly) assumed that propositions subsist independently of being thought. False propositions, then, were part of their ontology. Moore gave up the point in his 1910–11 lectures at Morley College London. Cf. Korhonen (2009).
30. In the *Tractatus logico-philosophicus* 4.061, Wittgenstein revives Frege's view, and in 4.0621 (b) he writes "The propositions 'p' and '~p' have opposite sense, but there corresponds to them one and the same reality." But a few lines later, in 4.063, he distances himself from some relevant details implied by Frege's (1893) ideography, details linked with the horizontal. Bonino (2008) discusses at length this section of the *Tractatus*.
31. We distinguish a thousand different colors, and in direct contrast millions.

32. Bonino (2008) in many places (in Part I.6, see, for instance, 70–71) criticizes a similar idea, which was, for instance, argued for by Demos (1917) against Russell. The problem, according to Bonino, is that this idea, which Demos calls 'opposition,' depends 'on a plurality of (accidental) features of the world' (71). It depends on features of the world, yes, but one does not forget accidental features to get rid of a problem. Wittgenstein (1921) assumed the mutual logical independence of any elementary proposition, but he was wrong. If my t-shirt is green, it is not blue, and that's because if it has a color of a wavelength, it has not a color of any other wavelength.
33. Frege seems to miss the analogy in that he writes: "Every declarative sentence concerned with the referents of its words is therefore to be regarded as a proper name, and its referent, if it exists, is either the true or the false." (1892: 34 [1948: 216]).
    Austin (1950) (126, fn 2) traces a parallel between facts and existence, which is also a parallel between the semantics of sentences and noun phrases.
34. Instead, the sentences 'The author of the three volumes of *Principia Mathematica* is a set' and 'The author of *Principia philosophica* is a set' would turn out true. That is a technical issue, which I believe Russell noticed and made him qualify as artificial Frege's proposal. Even taking into account the more open-mindedness Frege exhibits in "On Sense and Meaning" concerning vacuous names in ordinary language talk, Frege's choice disconnects falsity from representation, thereby overlooking the fact that the false introduces us to pretending, an intriguing ado about nothing, and to an investigation of what is possible via what is not, and what is not because it cannot be.
35. See also 1884: x [1960: xxii]. Here, the term *proposition* would most likely today be substituted by *sentence*, but see Geach (1962: 51–2).
36. Dummett writes:

> If sentences are merely a special case of complex proper names, if the True and the False are merely two particular objects amid a universe of objects, then, after all, there is nothing unique about sentences: whatever was thought to be special about them should be ascribed, rather, to proper names – complete expressions – in general. This was the most disastrous of the effects of the misbegotten doctrine that sentences are a species of complex name, which dominated Frege's later period: to rob him of the insight that sentences play a unique role, and that the role of almost every other linguistic expression (every expression whose contribution to meaning falls within the division of sense) consists in its part in forming sentences. After the adoption of the new doctrine,

> only the ghost of the original thesis could remain: the sense of a word now had to be thought of as relating, not particularly to the determination of the truth-value of a sentence containing it, but, more generally, to the determination of the referent of a complex proper name containing it. (1973: 196).
> But see also Dummett (1973: 181).

37. Burge (1986: 97 [2005: 83]). Burge uses 'denotation' and not 'reference' as is done in the classical translation of Frege's papers. Whereas reference is not mediated, denotation is – in Frege, it is mediated by the sense of the expression.
38. Thus at least taking apart its referential use *à la* Donnellan.
39. E. Bach (1968) argues this to be the origin of any description, definite or indefinite.
40. Frege speaks of the name of a truth-value and adds that " — Δ is the True if Δ is the True" (the thought that Δ is true is true if and only if Δ is true). There are two problems here though. (i) Burge describes the statement as one of truth redundancy (1986: 127). I am not so sure, because in the horizontal there is something metalinguistic that the linguistic predicate 'true' misses. (ii) Frege does not place explicit constraints on the form of Δ. Can Δ be a sentence or also a description such as *the true*? Against the second option there is Frege's claim that the word *true* is senseless, as he argues much later in 1918. Here, I use *senseless* as opposed to *nonsense*, *à la* Wittgenstein.

References

Angelelli, I. 1978. Die Zweideutigkeit von Freges 'Sinn' und 'Bedeutung'. *Allgemeine Zeitschrift für Philosophie* 3: 62–66.

Austin, J.L. 1950. Truth. In *Proceedings of the Aristotelian Society*. Supp. Vol. 24, 111–129.

Bach, E. 1968. Nouns and Noun Phrases. In *Universals in Linguistic Theory*, ed. E. Bach and R.T. Harms, 91–122. New York: Holt, Rinehart and Winston.

Bell, D. 1980. On the Translation of Frege's *Bedeutung*. *Analysis* 40: 191–195.

Bernini, S. 2018. L'ottica fregeana degli enunciati come nomi. *Rivista di filosofia*.

Bonino, G. 2008. *The Arrow and the Point/Russell and Wittgenstein's Tractatus*. Frankfurt: Ontos Verlag.

Burge, T. 1984. The Concept of Truth in Frege's Program. *Philosophia Naturalis* 21: 507–512; repr. in Tyler Burge. 2005. *Truth, Thought, Reason/Essays on Frege*, 77–82. Oxford: Clarendon Press.

Burge, T. 1986. Frege on Truth. In *Frege Synthesized*, ed. Haaparanta & J. Hintikk, 97–154. Dordrecht: Reidel; repr. in Tyler Burge. 2005. *Truth, Thought, Reason/Essays on Frege*, 83–132. Oxford: Clarendon Press.

Burge, T. 2005. Postscript to Frege on Truth. In *Truth, Thought, Reason/Essays on Frege*, ed. Tyler Burge, 133–152. Oxford: Clarendon Press.

Cheney, C., and C. Pigden. 2006. Negative Truths from Positive Facts. *Australasian Journal of Philosophy* 84: 249–265.

Davidson, D. 1967. The Logical Form of Action Sentences. In *The Logic of Decision and Action*, ed. N. Rescher, 81–95. Pittsburgh: University of Pittsburgh Press.

Davidson, D. 1969. True to the Facts. *The Journal of Philosophy* 66: 748–764.

Demos, R. 1917. A Discussion of Certain Types of Negative Propositions. *Mind* 26: 188–196.

Ducasse, C.J. 1940. Propositions, Opinions, Sentences and Facts. *The Journal of Philosophy* 37: 701–711.

Dummett, M. 1973. *Frege/Philosophy of Language*. London: Duckworth.

Frege, G. 1879. *Begriffsschrift eine der arithmetischen nachgebildete Formelsprache der reinen Denkens*. Halle: Louis Nebert; Eng. Tr. In *Begriffsschrift, a Formula Language, Modeled Upon That of Arithmetic, for Pure Thought*, in *From Frege to Gödel*, ed. J. van Heijenoort, 1–82. Cambridge, MA: Harvard University Press, 1967.

Frege, G. 1884. *Die Grundlagen der Arithmetik*. Breslau: Koebner.

Frege, G. 1891. Function und Begriff. address given to the *Jenaische Gesellschaft für Medicin und Naturwissenschaft*, January 9, 1891; Eng. tr. Function and concept. In G. Frege *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black, 21–41. Oxford: Blackwell, 1960.

Frege, G. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50; Eng. tr. On Sense and Reference. In G. Frege *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black, 56–78. Oxford: Blackwell, 1960.

Frege, G. 1893. *Grundgesetze der Arithmetik/Begriffsschrift abgeleitet*. Band I, Jena: Pohle; Eng. Tr. of Introduction and $$ 1–52, *Basic Laws of Arithmetic/Exposition of the System* with an introduction by M. Furth, Berkeley and Los Angeles: University of California Press, 1964.

Frege, G. 1918. Der Gedanke/Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus* 1: 58–77; Eng. tr. The Thought. *Mind* 65: 289–311, 1956.

Gabriel, G. 1984. Fregean Connection: *Bedeutung*, Value and Truth-Value. *Philosophical Quarterly* 34: 372–376.

Geach, P. 1962. *Reference and Generality*. Ithaca, NY and London: Cornell University Press.

Heck, R.G., and R. May. 2018. Truth in Frege. In *Oxford Handbook of Truth*, ed. M. Glanzberg. Oxford: Oxford University Press.

Holenstein, E. 1983. The Meaning of 'Bedeutung' in Frege. A Philological Approach. In *History of Semiotics*, ed. A. Eschbach and J. Trabant, 311–321. Amsterdam/Philadelphia: Benjamins.

Jubien, M. 1991. Could This Be Magic? *The Philosophical Review* 100: 247–267.
Korhonen, A. 2009. Russell's Early Metaphysics of Propositions. *Prolegomena* 8: 159–192.
Künne, W. 2003. *Conceptions of Truth*. Oxford: Clarendon Press.
Künne, W. 2008. Frege on Truths, Truth and the True. *Studia Philosophica Estonica* 1: 5–42.
Lebens, S. 2017. *Bertrand Russell and the Nature of Proposition*. New York and London: Routledge.
Lockwood, M. 1975. On Predicating Proper Names. *The Philosophical Review* 84: 471–498.
Long, P., and R. White. 1980. On the Translation of Frege's *Bedeutung*: A Reply to Dr. Bell. *Analysis* 40: 196–220.
Moro, A. 2017. *A Brief History of the Verb to Be*. Cambridge, MA: MIT Press.
Picardi, E. 1981. *Assertibility and Truth/A Study of Fregean Themes*. Bologna: CLUEB.
Picardi, E. 1994. *La chimica dei concetti/Linguaggio, logica, psicologia*. Bologna: Il Mulino.
Prior, A.N. 1971. *Objects of Thought*. Oxford: Clarendon Press.
Quine, W.v.O. 1974. *Methods of Logic*, 4th ed. Cambridge, MA: Harvard University Press.
Ricketts, T. 2003. Quantification, Sentences, and Truth-Values. *Manuscrito* 26: 389–424.
Russell, B. 1918–1919. The Philosophy of Logical Atomism. *The Monist* 28: 495–527; 29: 32–63, 190–222, 345–380.
Shramko, Y., and H. Wansing. 2005. Some Useful 16-Valued Logics: How a Computer Network Should Think. *Journal of Philosophical Logic* 34: 121–153.
Textor, M. 2011. *Frege on Sense and Reference*. London: Routledge.
Textor, M. 2016. States of Affairs. *Stanford Encyclopedia of Philosophy*, Winter 2016 Edition, https://plato.stanford.edu/archives/win2016/entries/states-of-affairs/.
Tugendhat, E. 1970. The Meaning of 'Bedeutung' in Frege. *Analysis* 30: 177–189.
Wittgenstein, L. 1921. Logisch-Philosophische Abhandlung. *Annalen der Naturphilosophie* 14: 185–262; Engl. tr. C.G. Odgen *Tractatus Logico-Philosophicus*. London: Kegan Paul, 1922.

Was Frege a Logicist for Arithmetic?

Marco Panza

1 Introduction

Was Frege a logicist? At first glance, the question has an obvious answer: Frege was not only a logicist; he was (possibly together with Dedekind, but in another sense)[1] the very founder of logicism, and remains (one of) the most emblematic representatives of it.

Eva would have possibly agreed. Still, supposing I had insisted in asking, her intellectual curiosity and love for philosophical and historical

The paper originated in a discussion during a conference at Keio Univ. (Tokyo), and preliminary versions of it were presented at the IHPST (Paris), Univ. San Raffaele (Milano), Chapman Univ. (Orange, CA), Univ. de Lorraine (Nancy), and the Czech Acad. of Sci., Dept. of Philophy (Prague). I thank the audience of all these talks for useful objections and suggestions that greatly helped me to improve the paper. Special thanks to J. Bertran-San Millán, F. Boccuni, A. Coliva, G. Heinzmann, R. May, A. Sereni, D. Struppa, G. Sundholm, J. Tappenden and P. Wagner.

M. Panza (✉)
CNRS, IHPST (CNRS and University of Paris 1, Panthéon-Sorbonne), Paris, France
marco.panza@univ-paris1.fr

M. Panza
Chapman University, Orange, CA, USA

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_5

discussion would have certainly made her reply with another question: 'How could you think that he wasn't?' Answering this other question in detail would have taken more than a discussion. Possibly, we would have decided to tackle the question brick by brick, which would have been the occasion of more than a dinner together. Space limitation forces me to offer her, here, only a quite summary version of what I would have said her during the first of these dinners.

In particular, I shall limit my attention to arithmetic, only. I hope to have room to expound these remarks and extend my scrutiny to real analysis in some other occasion.[2] On this matter, let me only say that what I shall say about the former seems to me to nicely fit with what Andrea Sereni (a friend which I met thanks to Eva) and I have argued concerning Frege's conception of the application constraint both for natural and real numbers (Panza and Sereni, Forthcoming). Taken together, with those advanced in this last paper, my following considerations are aimed to offer a quite dissident picture of Frege's foundational purpose, depicting him more as a mathematician aiming at providing mathematics with an appropriate architecture, than as a philosopher aiming at securing its epistemic grounds. It is this picture that I hope to be able to refine and complete elsewhere.

My present claim is that Frege's primary foundational purpose concerning arithmetic was neither that of making natural numbers logical objects, nor that of making arithmetic a part of logic, but rather that of assigning to it an appropriate place in the architectonics of mathematics and knowledge.

2 A Terminological Remark

Let me begin with an obvious but relevant terminological remark. Frege never termed himself a logicist, and never used any German word for 'logicism' and its cognates.

In its contemporary sense, such a term appeared quite late (Grattan-Guinness 2000, pp. 479, 501). It was almost simultaneously and independently firstly used by Fraenkel in his *Einleitung in die Mengenlehre* (1928) and by Carnap, in his *Abriss der Logistik* (1929). The former used 'logistische Schule' and 'Logizismus', to refer to Russell and Whitehead's views, as opposed to those of "intuitionists" and "formalists" (Fraenkel 1928, p. 263). The latter used 'logizistich' and 'Lozicism' to characterize "a philosophical direction with a strong or even excessive emphasis on the logical point of view" (Carnap 1929, p. 3).

On September 5th, 1930, Carnap, Heyting and von Neumann delivered their well-known talks about the "*logizstische*", "*intuitionistische*" and

"*formalistische*" foundation of mathematics (Carnap 1931; Heying 1931; von Neumann 1931). Carnap's lecture definitively codified the use of 'logicism [*Logizismus*]' in our present sense: forty-four years after the publication of *Grundlagen* (Frege 1884), and twenty-seven after that of second volume of *Grundgesetze* (Frege 1893–1903).

Following Fraenkel (rather than his own earlier use), Carnap called thus "the thesis that mathematics is reducible to logic", is "nothing but a part of logic", and assigned to Russell the role of "chief proponent" of it, while adding that "Frege was the first to espouse this view" (Carnap 1931, p. 91; PMBP, p. 31). This is openly false, if mathematics is intended to include geometry, as it should be, in Russell's view. Hence, if Frege was "the first to espouse" it, the view cannot be but that a piece of mathematics is reducible to, or is a part of logic. But which piece? Arithmetic is the only candidate I'll consider here.

3 Some (Dissonant) Quotes

Frege's argument about arithmetic goes in three stages: in *Grundlagen*, alternative conceptions are criticized; in *Grundlagen*, again, *Anzahlen* are informally identified with extensions of second-level concepts, and natural numbers are singled out among them[3]; in *Grundgesetze*, this informal treatment is turned into a formal definition. What is the point Frege aimed to make, by following this tripartite strategy? Different passages drawn both from *Grundlagen* and *Grundgesetze* suggest different responses.

Here are some from *Grundlagen* (Frege 1884, *Einleitung*, pp. IV, XI, §§87; GLAA, pp. xvi, xxi, 99, 102)[4]:

> The present work will make it clear that even an inference like that from n to n+1, which on the face of it is peculiar to mathematics, is based on the general laws of logic.
>
> I have felt bound to go back […] further into the general logical foundations [of mathematics][…].
>
> I hope to have made it plausible that the laws of arithmetic are analytic judgements[5] and consequently a priori. Arithmetic thus becomes nothing but further pursued logic, and every arithmetic statement becomes a law of logic.
>
> I do not claim to have made the analytic character of arithmetical statements more than plausible, because it can still always be doubted whether their proof only proceeds from purely logical laws […]. This misgiving […] can only be removed by producing a gapless chain of inferences […].

Part II of *Grundgesetze* is intended to achieve this last task. Here is how Frege announces it (1893–1903, *Einleitung*, vol. I, p. 1; GGAER, p. 1_1):

> In my *Grundlagen der Arithmetik* […] I aimed to make it plausible that arithmetic is a branch of logic […]. In the present book this is now to be established by deduction of the simplest laws of *Anzahlen* by logical means alone.

Some lines later, Frege terms this view his 'thought' (1893–1903, *Einleitung*, vol. I, p. 3; GGAER, p. 3_1):

> If my thought, that arithmetic is a branch of pure logic, is correct, then […].

This confirms what he implies in the Foreword (Frege 1893–1903, *Vorwort*, vol. I, p. VIII; GGAER, p. $VIII_1$):

> Mr Dedekind too is of the opinion that the theory of numbers is a part of logic.

In none of these passages, the positive claim is made that arithmetic is a part of logic. Frege maintains to have made it plausible, and to have designed his proofs in order to support it; but he does not take it as an established fact; he merely advances it as a "thought" or "opinion". Still the source of uncertainty is identified. It pertains to Basic Law V (BLV, from now on: Frege 1893–1903, *Vorwort*, vol. I, p. VII; GGAER, p. VII_1):

> […], a dispute can arise only concerning my basic law of value-ranges (V) […]. I take it to be purely logical. At any rate, the place is hereby marked where there has to be a decision.

These passages, coming from *Grundlagen* and the first volume of *Grundgesetze*, suggest that Frege certainly endorsed that arithmetic is a part of logic, but admitted to have not established it. What he considered to have established was only that this is so, if BLV is a logical law. Still, in this first volume, he makes no effort to argue that BLV is actually a logical law. He takes it to be so, but offers no real argument for it. This suggests that his purpose was less to establish that arithmetic is a part of

logic than to ascertain that it belongs to the same field as BLV, namely it is part of a more general theory of concepts and their extensions.

Things seem to change in the second volume. Though he never explicitly argues, yet, that arithmetic is a part of logic, he has no reticence, there, in claiming that *Anzahlen* are logical objects. While discussing Cantor's views, he distinguishes "physical […] [from] logical objects", and advances that the latter "include our *Anzahlen*" (Frege 1893–1903, §II.74; GGAER, p. 86_2). Later, while discussing Dedekind's views, he observes that (Frege 1893–1903, §II.147; GGAER, p. 149_2):

> If there are logical objects at all—and the objects of arithmetic are such—then there must also be a means to grasp them, to recognize them. […] [BLV] serves for this purpose. Without such a means, a scientific foundation of arithmetic would be impossible.

After having become aware of Russell's paradox, he maintained the same view (Frege 1893–1903, *Nachwort*, vol. II, pp. 253 and 265; GGAER, pp. 253_2 and 265_2):

> I have never concealed […] that […][BLV] is not as obvious as the others, nor as must properly be required of a logical law. […]. I would gladly have dispensed with this foundation if I had known of some substitute for it. Even now, I do not see how arithmetic can be founded scientifically, how the numbers can be apprehended as logical objects […], if it is not—at least conditionally—permissible to pass from a concept to its extension. […] This question may be viewed as the fundamental problem of arithmetic: how are we to apprehend logical objects, in particular, the numbers? What justifies us to acknowledge numbers as objects?

The end of this passage intimates that Frege was overall concerned with numbers being objects, and making these objects logical was a means for arguing for that, without involving intuition. The beginning suggests this is so also true for extensions: rather than aiming at making arithmetic ensue from BLV, he seems here to take extensions as the most obvious objects to identify with numbers for making them objects.

Is this a shift occurring after the reception of Russell's letter? Other passages from *Grundlagen* suggest that this is not so. Better, that no shift occurred, in fact; that Frege was constantly oscillating among different perspectives. Here is one (1884, §107; GLAA, p. 117):

> […] the sense of the expression 'extension of a concept' is taken as already known. This way of getting over the difficulty cannot be expected to meet with universal approval […]. By the way, I attach no decisive importance to the use of extensions of concepts.

Coming back to *Grundgesetze* in the light of this claim and comparing it with the previous passages make one think that Frege considered both extensions and BLV as convenient, but replaceable, tools to define *Anzahlen*. Yes, but with which purpose?

Before this passage, no mention is made of logic. Frege discusses (1884, §§104–105; GLAA, p. 114–115) the possibility of extending his views on *Anzahlen* to "other numbers", and observes that also for them,

> everything will in the end come down to the search of a judgement-content, which can be transformed into an identity whose sides precisely are the new numbers. […] If we proceed as […][for *Anzahlen*], then the new numbers will be given to us as extensions of concepts […][, as], objects which we do not come to know as something alien coming from outside, through the medium of the senses, but with objects given directly to our reason and, as its nearest kin, utterly transparent to it.

What Frege seems here to grant is less having made arithmetic a part of logic, than having made *Anzahlen* "given directly to our reason […], utterly transparent to it", rather than "known as something alien coming from outside, through the medium of senses".

In Kantian language, this is the same as having made knowledge of *Anzahlen* and judgements about them a priori. Frege's notion of a priori is not coincident with Kant's. For him, a truth (rather than a judgment) is a priori if "its proof proceeds as a whole from general laws, which neither need nor admit of proof" (Frege 1884, §3; GGAA, p. 4). I have no room here to discuss the differences among the two notions. What is relevant is that also for Frege himself his achievement might well be outlined in term of apriority as that of having made truths about *Anzahlen*, and, consequently, arithmetical truths a priori. It is then relevant to observe that Frege's explanation of apriority makes logical laws a priori, but not a priori truths logical. Moreover, Frege's doubts about the logicality, but not the truth of BLV, suggest that he admitted non-logical a priori, as it is also confirmed by the following passage, already quoted in part:

> We can distinguish physical from logical objects, by which […] no exhaustive classification is […] given.

Hence, making truths about *Anzahlen* and/or knowledge of them a priori was possibly not the same, for Frege, as making them logical objects.

* * *

Far from being concordant, when taken all together, the previous passages suggest three different possible understandings of Frege's purpose.

According to the first understanding, Frege's primary aim was to make arithmetic a part of a general theory of extensions of concepts.

According to the two others, extensions should be rather seen as nothing but a convenient tool to be used for reaching an independent purpose.

According to the second understainding, this purpose was that currently ascribed to the logicist Frege, i.e. to make *Anzahlen* logical objects and arithmetic a part of logic.

According to the third understanding, this propose was rather that of making truths about *Anzahlen*, and/or knowledge of them a priori, namely transparent to reason without the medium of senses and intuition.

This variety of understandings invites us to eschew general claims, and to search rather an answer to our question within Frege's definitions of *Anzahlen* and, among them, of natural numbers.

4 Concepts and Extensions

It is enough to parse, even quite cursorily (as I shall do below, in §5), the (informal) definition of natural numbers Frege offers in *Grundlagen* to realize that his treatment of arithmetic could have left extensions aside, if Hume Principle (HP, from now on) had been independently admitted. The (formal) definition of *Grundgesetze* makes this a little bit less evident (because of the technical apparatus used to avoid second-order quantification as much as possible, which makes value-ranges appear at face value virtually in any definition and theorem), but could certainly not have concealed it to Frege's own eyes. There is, then, no doubt that

he was perfectly aware of this possibility. But, then, why did he not only adopt BLV and give a so crucial role to value-ranges and, particularly, extensions, but even refrain from giving them up after having become aware of Russell's paradox?

Appealing to Caesar problem does not set the question. Since, taken as such, Caesar problem also arises for BLV. Another answer comes from an argument Frege advances, in the second volume of *Grundgesetze* to deny that BLV results in a creative definition (Frege 1893–1903, §§II.146–147; GGAER, pp. 147_2–148_2):

> […] it could be pointed out that […] we ourselves created new objects, namely value-ranges. […] We did not list properties and say: we create a thing that has them. Rather, we said: if one function […] and a second function are so constituted that both always have the same value for the same argument, then one may say instead: the value-range of the first function is the same as the that of the second. We recognize something in common to both functions and this we call the value-range both of the first and of the second function. That we have the right to the acknowledging of what is common, and that, accordingly, we can convert the generality of an equality into an equality (identity), must be regarded as a basic law of logic.

Hence, for Frege, passing from recognizing something in common to two functions to asserting an identity is licensed by a "basic law of logic". If this were admitted and BLV merely licensed this passage, it would, then, be a logical law. At the best of my knowledge, this is the only substantial argument Frege ever advanced for the logicality of BLV. But, so conceived, it is openly unsound. Since BLV does much more than this. It licenses passing from recognizing that two functions bear to each other an equivalence relation, to associating one same object to both of them, and, vice versa, from recognizing—or, better, admitting—that two functions are associated to the same object in the same canonical way, to concluding that they bear to each other this same equivalence relation.

This makes the crucial point appear. This is not so much that Frege does not consider the left-to-right implication involved in BLV (an allegation that, by the way, would be rightful only if it were admitted that he actually considered this argument as an argument in favor to the logicality of BLV). Rather it is that he seems to take values-ranges to be

intrinsically and openly there within functions, and, in particular, extensions to be intrinsically and openly there within concepts. This makes BLV merely make explicit a necessary and sufficient identity condition for value-ranges of first-level one-argument functions (a condition that, in the previous argument, is in fact only regarded as sufficient). If it were so, BLV would essentially differ from HP. Since, whereas values-ranges would be intrinsically and openly there within functions, would come directly together with them, *Anzahlen* certainly do not come directly together with first-level concepts. In §§5 and 6, below, I shall shortly account for Frege's definitions of *Anzahlen* in *Grundlagen* and *Grundgesetze*. For the time being, only some clues are in order. According to both definitions, whatever *Anzahl* is the *Anzahl* of a first-level concept. The former definition identifies it with the extension of the second-level concept of being (a first-level concept) equinumerous with this first-level concept. Hence, conforming to it, an *Anzahl* comes with the concept which it is the *Anzahl* of only insofar as this concept is associated to a second-level one, of which this *Anzahl* is the extension. The latter definition identifies the *Anzahl* of a certain first-level concept with the extension of the first-level concept of being the value-range of a first-level function which takes the True as its value for as many objects as argument as those which fall under this first-level concept. Hence, conforming to it, an *Anzahl* comes with the concept which it is the *Anzahl* of only insofar as this concept is associated to another first-level concept, namely with the concept of being the value-range of an appropriate function. It is, then, clear that, for Frege, HP does much more than making explicit a necessary and sufficient identity condition for something which is intrinsically and openly there within concepts, that comes directly together with them. By asserting that the same *Anzahl* is ascribed to two first-level concepts if and only if these concepts are equinumerous,[6] it ensures that the fact that these first-level concepts are associated to two other concepts in the same canonical way makes the extensions of these other concepts to be the same, and, then, to count as the common *Anzahl* of the former concepts if and only if this last condition obtains. Hence, HP requires a proof, whereas BLV "neither needs nor admits of" it.

But should we grant that values-ranges are intrinsically and openly there within functions, that they come directly together with them? If yes, and only if yes, BLV would have a chance of being a law of logic. This is openly questionable, however, and this is just where the residual

doubt about BLV's logicality seems to come from, for Frege. To settle this doubt, Frege should have elucidated the notion of a value-range. But this would have made him run the risk of explaining extensions through sets, thus making the latter come logically before the former: a view he was strongly opposed (Benis-Sinaceur et al. 2015, *Introduction* and Chapter 2).

In *Grundlagen*, Frege merely takes for granted "what the extension of a concept is" (Frege 1884, §69; GLAA, p. 80, footnote 1). In *Grundgesetze*, he recognizes that BLV offers no way to identify value-ranges (which seems, by the way, in contrast with his taking them to openly come directly together with functions), and merely makes a stipulation, identifying the extensions of two concepts under which only the True and the False respectively fall with the True and the False themselves (Frege 1893–1903, §I.10, GGAER, pp. 16_1–18_1). This makes his view unstable. Since not explaining what extensions are entails making impossible to explain what *Anzahlen* are. And merely stipulating that two value-ranges are the True and the False makes the nature of *Anzahlen* subject to stipulation.

What could have, then, made natural numbers logical objects? Identifying these numbers with some extensions of concepts would not have been enough, since logic does not encompass particular concepts; it merely deals with concepts in general. Hence, only the particular nature of the concepts these numbers were taken to be extensions of could have ensured this. To answer the question, we have, then, no other resources than look at these concepts, namely parse, even shortly, Frege's definitions of *Anzahlen*, and natural numbers, among them.

5 The Definition of *Grundlagen*

Let us begin with the definition of *Grundlagen* and, for sake of simplicity, render it through a compact notation (still remembering that it is informal, so that this notation is nothing but a suitable tool to abbreviate sentences written in natural, though codified language).

Let P and Q be whatever first-level concepts, $\mathscr{P}$ whatever second-level concept, and 'X' and 'x' variables ranging on first-level concepts and objects, respectively. Read: '$\#P$' as 'the *Anzahl* of P'; '$\varepsilon\mathscr{P}$' as 'the extension of $\mathscr{P}$', '$P \approx Q$' as 'P is equinumerous with Q', and '$\mathbf{N}_{\mathbf{A}}x$' as 'x is an *Anzahl*'. Then, Frege (1884, §§68 and 72) begins by stating that:

$$\text{GL.}i) \quad \#P =_{df} \mathcal{E}[X : X \approx P]$$

$$\text{GL.}ii) \quad \forall x[\mathbf{N_A}x \Leftrightarrow \exists X[x = \#X]]$$

This is enough to make him able to informally prove HP (ibidem, §73). After having proved it (ibidem, §§74 and 76), Frege states that:

$$\text{GL.}iii) \quad 0 =_{df} \#[x : x \neq x]$$

and

$$\text{GL.}iv) \quad \forall x, y\big[x\mathbf{F_I}y \Leftrightarrow \exists X \exists z\big[x = \#X \wedge Xz \wedge y = \#[w : Xw \wedge w \neq z]\big]\big]$$

where 'y', 'z' and 'w' are other variables ranging on objects, and '$x\mathbf{F_I}y$' is to be read as 'x follows immediately after y in the natural sequence of numbers'.[7] This allows him to observe (ibidem, §77) that from HP it follows that

$$\#[x : x = 0]\mathbf{F_I}0$$

which suggests stating that

$$\text{GL.}v) \quad 1 =_{df} \#[x : x = 0]$$

Insofar as from (GL.*iv*) and (GL.*v*) it also follows that $\#[x : x = 0 \vee x = 1]\mathbf{F_I}1$, Frege might have continued by stating that $2 =_{df} \#[x : x = 0 \vee x = 1]$, etc. Still, this would have not allowed him to explicitly define natural numbers in general, but only to recursively define indefinitely many such numbers. Hence, by leaving (GL.*v*) aside, he rather defines the new first-level binary relation of following after in the natural sequence of numbers—which I shall shortly designate by '**F**'—as the the strong ancestral of $\mathbf{F_I}$ (ibidem, 79 and 81), namely:

$$\text{GL.}vi) \quad \forall x, y\big[x\mathbf{F}y \Leftrightarrow \forall X\big[\big[\forall z\big[z\mathbf{F_I}y \Rightarrow Xz\big] \wedge \forall z, w[(Xz \wedge w\mathbf{F_I}z) \Rightarrow Xw]\big] \Rightarrow Xx\big]\big],$$

where '$x\mathbf{F}y$' is to be read as 'x follows after y in the natural sequence of numbers'. Then, based on this relation, he defines (ibidem, §81) another first-level binary relation, to be conceived as the relation of belonging to the natural sequence of numbers either beginning or ending with, the difference of the two relations merely depending on the order of relata. The definition is this:

$$\text{GL.}vii) \quad \forall x, y[x\mathbf{S}y \Leftrightarrow (x\mathbf{F}y \vee x = y)]$$

where '$x\mathbf{S}y$' is to be read either as 'x belongs to the natural sequence of numbers beginning with x' or 'y belongs to natural sequence of numbers ending with x'.

Finally (Frege 1884, §§79 and 82–83; Boolos and Heck 1998), he sketches a proof of

$$\forall x[\mathcal{C}(x) \Rightarrow \#[z : x\mathbf{S}z]\mathbf{F}_{\mathbf{I}}x]$$

where: '$\mathcal{C}(x)$' stands from 'x meets a condition to be specified', by so showing that this holds if

$$\forall x[\mathcal{C}(x) \Leftrightarrow x\mathbf{S}0]$$

which suggests abbreviating this condition, namely the condition that x belongs to a succession beginning with 0, with 'x is a finite *Anzahl*', or, by short, '$\mathbf{N}x$', from which it follows that

$$\text{GL.}viii) \quad \forall x[\mathbf{N}x \Leftrightarrow x\mathbf{S}0]$$

and

$$\forall x[\mathbf{N}x \Rightarrow \exists y[\mathbf{N}_{\mathbf{A}}y \wedge y\mathbf{F}_{\mathbf{I}}x]] \qquad (\text{SUCC}_{\text{GL}})$$

What is crucial for us, namely definition (GL.*viii*), is presented as quite marginal by Frege: he does not only merely qualify it as a "convenient abbreviation", but he comes to it while establishing a sufficient condition for it to hold that an *Anzahl* be immediately followed by another *Anzahl*. In contemporary terminology: Frege recognizes that only certain *Anzahlen* meet Peano's successor axiom, and identify finite ones—or natural numbers, as we currently call them—as some of those which do; but he seems much more interested in enquiring about the properties of *Anzahlen* taken as such, than in singling out natural numbers among them. This is all the more clear that, after having established this abbreviation, he quickly leaves natural numbers aside to deal with *Endloss*: the *Anzahl* of the concept of being a finite *Anzahl*.

His main point seems, then, that of establishing that *Anzahlen* are numbers of first-level concepts—which allows studying them within a general theory of concepts and objects—and that some of them satisfy the fundamental theorems of arithmetic. His purpose, in other terms, seems more that of immersing arithmetic within a theory of numbers of first-level concepts, than that of developing it there.

This being said, some remarks on definitions (GL.*i*–*viii*) are in order. The first one is that extensions explicitly appear only in the first of these

definitions and in the proof of HP. This suggests that Frege was considering them useful only to make this proof possible, so as to show that *Anzahlen* come together with fist-level concepts, though not directly.

The adverb 'only' in this claim is not to be intended as aiming to undermine the role of extensions. To see how crucial this role is, notice that (SUCC_{GL}) is all that Frege offers in *Grundlagen* concerning the existence of *Anzahlen*. On the standard interpretation of the existential quantifier, this theorem asserts that natural numbers exist, if 0 does. Still, granting this interpretation is not enough, yet, to allow the proof of this theorem to establish this. To this purpose, it is also required to grant that Frege's definitions have a model, that is, that there are objects complying with them. From this, it also follows that 0 exists, but this condition makes it flagrantly circular to appeal to this theorem to conclude that natural numbers exist.

Though Frege would have certainly not reasoned in terms of models, he could have not ignored that alleging that his proof of (SUCC_{GL}) provides an argument for the existence of natural numbers would have been circular. This suggests that he was taking the question of the existence, not only of natural numbers, but, more generally, of *Anzahlen*, settled by his definition of them as extensions of openly identified concepts: he was possibly taking *Anzahlen* to exist, just in virtue of the fact that these concepts are given, and their extensions come directly with them.

Still, Russell's paradox apart—that is, also admitting to adopt an appropriate variant of Frege's framework, where the relevant extensions could be consistently handled[8]—taking *Anzahlen* to be identified with extensions (of second-level) concepts has a price, which is often unnoticed.

If the general nature of *Anzahlen* merely depended on HP and definition (GL.*ii*), one could think that the particular nature of the *Anzahl* of a certain concept only depends on the nature of this concept, namely that the particular nature of $\#P$ only depends on the nature of P. If this were granted, it would be enough to admit that this last concept is logical, to conclude that $\#P$ is either a logical object, provided that HP were taken to be logical, or an object having with logic the same sort of relation as HP. But if $\#P$ is identified with the extension of the second-level concept $[X : X \approx P]$, things change. Since, then, taking $\#P$ as a logical object just means taking this last extension as a logical object, and this is highly questionable, indeed. The reason is obvious: it seems hard to deny that the nature of this extension depends on which concepts are

equinumerous with P, which does not seem to be a question of logic, even if P is a concept like $[x : x \neq x]$ or $\left[x : \bigvee_{i=0}^{i=n} x = i\right]$ $(n = 0, 1, \ldots)$, which are the only relevant ones for deciding whether natural numbers are logical objects, and might be plausibly taken as logical ones (if the latter are recursively defined, of course).

There are two obvious ways to resist this objection: (*i*) either admitting that the nature of $\mathcal{E}[X : X \approx P]$, or at least its logicality, is not affected, after all, by which concepts are equinumerous with P; (*ii*) or that, among all concepts equinumerous with $[x : x \neq x]$ or $\left[x : \bigvee_{i=0}^{i=n} x = i\right]$, the only relevant ones for this matter are logical ones. The problem is that either option seems unavailable to Frege.

Against (*ii*) does not only militate the fact that, in the informal framework of *Grundlagen*, it is hard to establish which concepts are logical, but also, and overall, the obvious circularity of deciding whether $\mathcal{E}[X : X \approx [x : x \neq x]]$, or $\mathcal{E}\left[X : X \approx \left[x : \bigvee_{i=0}^{i=n} x = i\right]\right]$ are logical objects, based of this admission.

Against (*i*) militates the very conception of the extension of a concept coming from the tradition that Frege seems to rely on when taking as "known what the extension of a concept is", for example, Kant's definition of it as a "*sphaera*" which "rises up from the multitude of things that are contained under the concept" (Kant WL, p. 911; LLY, p. 354). If straightforwardly applied to second-level concepts, according to an intensional notion of a first-level concept, this conception openly suggests that the extensions of $[X : X \approx [x : x \neq x]]$ and $\left[X : X \approx \left[x : \bigvee_{i=0}^{i=n} x = i\right]\right]$ actually depend on which first-level concepts (intensionally speaking) fall under these concepts, and, consequently, on how many objects fall under whatever finite sortal first-level concept. It suggests, for example, that the extension of $\left[X : X \approx \left[x : \bigvee_{i=0}^{i=13} x = i\right]\right]$ depends on how many objects actually fall under the concept of eight-thousanders on Earth. In our actual word, there are just fourteen eight-thousanders on Earth, which makes the latter concept fall under the former. It is obvious, however, that this is not a matter of logic. Thus, the extension of the former concept can hardly be taken to be a logical object.

Many countermoves are possible against this conclusion.

The most obvious one consists in observing that nothing forces us to take Frege's conception of extensions of second-order concepts to be in agreement with Kant's view. This is, however, what is openly implied by Frege's own informal proof of HP (Frege 1884, §73). Since, in conducting it, he supposes that

$$\mathcal{E}[X : X \approx P] = \mathcal{E}[X : X \approx Q] \Leftrightarrow \forall X[X \approx P \Leftrightarrow X \approx Q]$$

for whatever first-level concepts P and Q, which suggests that he was just taking the extensions of $[X : X \approx P]$ and $[X : X \approx Q]$ to depend on which first-level concepts are equinumerous with P and Q.

Another countermove consists in admitting this last point, while denying that concepts are to be understood intensionally. Indeed, according to an extensional notion of a concept, the extension of $\left[X : X \approx \left[x : \bigvee_{i=0}^{i=13} x = i\right]\right]$ does not depend at all, for example, on how many objects fall under the concept of eight-thousanders on Earth. For what identifies this last concept is not the condition of being an eight-thousander on Earth, but rather the fourteen mountains that fall under it, taken as such, which would form the very same concept, invariably falling under $[X : X \approx \left[X : X \approx \left[x : \bigvee_{i=0}^{i=13} x = i\right]\right]$, even if the eight-thousander on Earth were other than them. In other words, on this conception, the extension of $[X : X \approx [x : x \neq x]]$ is merely depending (in one way or another) on a single first-level concept, namely the empty one, while that of $\left[X : X \approx \left[x : \bigvee_{i=0}^{i=h} x = i\right]\right]$, for whatever natural number h, merely depends on the totality of $(h + 1)$-uples, no matter how each of them might be intensionally conceived. In my view, this is not, and could not have been Frege's conception of a concept,[9] but, this apart, the point, here, is that ascribing it to him does not put him in a better position for admitting that the extensions of concepts like $\left[X : X \approx \left[x : \bigvee_{i=0}^{i=n} x = i\right]\right]$ $(n = 0, 1, \ldots)$, are logical objects. Since, even according to this conception, these extensions depend on how the world is, namely on which objects exist (though they do not depend on the way these objects are conceptually classified).

This second countermove suggests many other ways to dismiss the point, by appropriately extending or restricting the relevant first-order

domain, or, at least, by making it independent of the objects which actually exist. A similar strategy, directly applied, instead, to the second-order domain is suggested by option (*ii*), above. The difficulty here is not only that of imagining how to do it without making the argument in favor of the logicality of $\mathcal{E}[X : X \approx [x : x \neq x]]$ and $\mathcal{E}\left[X : X \approx \left[x : \bigvee_{i=0}^{i=n} x = i\right]\right]$ circular. It is also, and overall, that any similar move would be openly incompatible with Frege's universalist and realist conceptions.

Hence, granted these conceptions, it seems plain that Frege could have hardly admitted that the extensions he identifies with natural numbers are logical objects, simply because their particular nature depends on matters on which logic cannot be taken to have any sort of jurisdiction.

Arguably, senses and intuition are necessary to grasp this particular nature. Still, neither Frege's proof of HP and the theorems of arithmetic, nor his argument for the existence of natural numbers depend on grasping this nature. The previous definitions are sufficient for this. One might, then, think that Frege considered these definitions enough for making the relevant properties of natural numbers transparent to reason without the medium of senses or intuition. For, though they make the nature of the objects they define ungraspable without senses and intuition, they openly identify the concepts which these numbers are the extensions of, and allow to prove HP and these theorems. And so much the worse for logical objects, if they make natural numbers hardly be such objects.

This supports the third of the options listed at the end of §3, and is compatible with the first: it suggests that Frege's primary purpose concerning arithmetic was that of making truths about *Anzahlen*, and/or knowledge of them a priori, namely of showing that these numbers can (or have to) be conceived as objects whose relevant properties, though not their particular nature, are transparent to reason without the medium of senses and intuition, and of doing it by also locating arithmetics within a general theory of concepts and their numbers (where the latter are seen as extensions of the former).

This picture is quite different from that conveyed by the usual description of Frege as a logicist about arithmetic, and is rather close to assigning to him a quasi-structuralist conception. But it seems to be the one which his definition of *Grundlagen* more plausibly suggests.

6 The Definition of *Grundgesetze*

Mutatis mutandis, this is also the picture suggested by the definition of *Grundgesetze*. Contrary to that of *Grundlagen*, this is a formal definition, that is, it is supplied within a formal system. This makes a crucial difference, which is relevant in many important respects. Still, this system is formal in a quite different sense than in our modern one: on the one hand, it does require no interpretation, since it already comes with a fixed meaning for its symbols, though not with a model, properly speaking; on the other hand, its variables are supposed to vary on given, universal domains. We know today that it could in no way have a model, since it is inconsistent. This is not the important point here, however. This is not only (and not mainly) because we might always suppose to adopt a consistent variant of Frege's system,[10] but overall because the point here is not what Frege's (original) definition actually defines, but what he could have taken it to define. Hence, what is important here is, rather, that Frege thought that his system had a fixed interpretation or, at most, a family of fixed interpretations[11] (which we could regard today as the intended one, or ones). The question is, then, what sort of objects, natural numbers should, or better could, be taken to be according to Frege's definition and to this interpretation or these interpretations.

In the new formal setting, value-ranges are governed by BLV, which is restricted to first-level functions. This makes *ipso-facto* hopeless to define *Anzahlen* as extensions of second-level concepts. Hence, Frege rephrases his definitions, by replacing higher-level concepts with value-ranges of appropriate first-level functions. Once again, it is in order here to merely provide a short outline of his definition.

For this, I will rephrase the basic components of it by using the same notation used above for the definition of *Grundlagen*, appropriately modified and extended. It is important to notice, however, that this notation is no more to be intended as apt to shorten informal discursive statements, but rather statements (i.e. explicit definitions and theorems) belonging to a formal (deductive) system.

Let Φ and Ψ be whatever first-level one-argument and two-arguments function, respectively, and $\top$ the True (a designed object, for Frege, just as the False, $\bot$). To make a long story short,[12] let us denote with '$\mathfrak{C}_\Phi$' the (first-level) concept ARGUMENT FOR WHICH Φ TAKES THE VALUE $\top$,

and with '$\mathfrak{R}_\Psi$' the (first-level) relation: ORDERED PAIR OF ARGUMENTS FOR WHICH Ψ TAKES THE VALUE $\top$, namely:

$$\mathfrak{C}_\Phi =_{abbr.} [x : \Phi(x) = \top],$$

$$\mathfrak{R}_\Psi =_{abbr.} [x, y : \Psi(x, y) = \top],$$

where 'x' and 'y' range on objects. Hence, if Φ is a concept and Ψ a (binary) relation, $\mathfrak{C}_\Phi$ and $\mathfrak{R}_\Psi$ come respectively to coincide with Φ and Ψ. Still, as we shall see, it is important to observe that Frege does not confine himself to consider concepts and relations, but rather extends his concern to functions in general.

This is already clear by the new definition replacing (GL.*i*). Instead of defining the *Anzahl* of a first-level concept, he now defines the *Anzahl* of an object (Frege 1893–1903, §I.40). For whatever object a, let $\mathfrak{T}_a$ be, for short, the (first-level) concept ARGUMENT FOR WHICH THE FIRST-LEVEL ONE-ARGUMENT FUNCTION (IF ANY) OF WHICH a IS THE THE VALUE-RANGE TAKES THE VALUE $\top$, namely:

$$\mathfrak{T}_a =_{abbr.} [x : \exists\varphi[a = \mathcal{E}\varphi \wedge \mathfrak{C}_\Phi x]],$$

where 'φ' ranges on first-level one-argument functions, and, for whatever such function Φ, '$\mathcal{E}\Phi$' denotes the value-range of Φ. This makes '$\mathfrak{T}_{\mathcal{E}\Phi}$' designate the same concept as '$\mathfrak{C}_\Phi$'. Frege's new definition can, then, be rendered as follows:

$$\text{GG.}i) \quad \#a =_{df} \mathcal{E}[x : \mathfrak{T}_x \approx \mathfrak{T}_a].$$

The fact that '#' applies now to objects, rather than to first-level concepts is, technically speaking, less relevant than it might appear (we shall see later what this entails on an interpretative level). For short, dub 'plain' an object, if any, which is not the value-range of a first-level one-argument function. If a is not plain, then $\#a = \#\mathcal{E}\Phi$ for some such function Φ, so that '#' combines with '$\mathcal{E}$' by yielding the operator '$\#\mathcal{E}$', applying to this function. But what happens, instead, if a is plain? The first thing to be observed on this matter is that nothing, in Frege's system, can ensure that there are plain objects. Though it does not entail it, the stipulation Frege advances in I.10 rather suggests the contrary. But, even if it were supposed that there be such objects and a were one of them, the concept $\mathfrak{T}_a$ would reduce to $[x : x \neq x]$. Hence, independently whether a is plain or not, definition (GG.*i*) states that $\#a$ is the extension of the concept of being the value-range of a first-level one-argument function Φ such that $\mathfrak{C}_\Phi$ is equinumerous with $\mathfrak{T}_a$, which allows taking $\#a$ as the *Anzahl* of $\mathfrak{T}_a$.[13]

Frege's further definitions can, then, reduce to formal rephrasings of those of *Grundlagen*, under the replacement of first-level concepts with value-ranges of first-level one-argument functions.

First come these four definitions (Frege 1893–1903), §§I.42–43):

$$\text{GG.}ii) \quad \forall x[\mathbf{N_A}x \Leftrightarrow \exists z[x = \#z]],$$

$$\text{GG.}iii) \quad 0 =_{df} \#\mathcal{E}[x : x \neq x],$$

$$\text{GL.}iv) \quad \mathbf{f_I} =_{df} \mathcal{E}[x, y : \exists\varphi\exists z\left[\#\mathcal{E}[w : \mathfrak{C}_\varphi w \wedge w \neq z\right] = x \wedge \mathfrak{C}_\varphi z \wedge \#\mathcal{E}\varphi = y]],$$

$$\text{GG.}v) \quad 1 =_{df} \#\mathcal{E}[x : x = 0],$$

where, for whatever first-level two-arguments function Ψ, '$\mathcal{E}\Psi$' denotes the value-range of Ψ, and '$\mathbf{f_I}$' is intended to denote the extension of the relation of being immediately followed by "in the sequence of *Anzahlen*".[14]

Once definition (GG.*iv*) stated, Frege can easily define the extensions $\overset{\prime}{-}\mathbf{f_I}$ and $\overset{\prime}{\smile}\mathbf{f_I}$ of the strong and week ancestral of this last relation (1893–1903, §§I.45–46):

$$\text{GG.}vi) \quad \overset{\prime}{-}\mathbf{f_I} =_{df} \mathcal{E}\left[x, y : \exists\psi\left[\mathcal{E}\psi = \mathbf{f_I} \wedge \forall\varphi\left[\left(\begin{array}{l}\forall z[x\mathfrak{R}_\psi z \Rightarrow \mathfrak{C}_\varphi z]\wedge \\ \forall z, w[(\mathfrak{C}_\varphi z \wedge z\mathfrak{R}_\psi w) \Rightarrow \mathfrak{C}_\varphi w]\end{array}\right) \Rightarrow \mathfrak{C}_\varphi y\right]\right]\right],$$

$$\text{GG.}vii) \quad \overset{\prime}{\smile}\mathbf{f_I} =_{df} \mathcal{E}[x, y : \exists\psi[\overset{\prime}{-}\mathbf{f_I} = \mathcal{E}\psi \wedge (x\mathfrak{R}_\Psi y \vee x = y)]],$$

where 'ψ' ranges on first-level two-argument functions. Provided that $\mathbf{S}^{-1}$ be the (first-level binary) relation whose extension is $\overset{\prime}{\smile}\mathbf{f_I}$, namely the relation of belonging to the sequence of *Anzahlen* ending with, definition (GG*vii*) makes '$0\mathfrak{R}_{\mathbf{S}^{-1}}a$' stand for '$0\mathbf{S}^{-1}a = \top$', whatever object a might be. With the help of this definition, it is, then, finally easy to define the property of being a "finite *Anzahl*", i.e. a natural number (Frege 1893–1903, §I.46):

$$\text{GG.}viii) \quad \forall x[\mathbf{N}x \Leftrightarrow \exists\psi(\mathcal{E}\psi = \overset{\prime}{\smile}\mathbf{f_I} \wedge 0\mathfrak{R}_\psi x)].$$

Mutatis mutandis, what we have noticed above for definition (GL*viii*) also holds for definition (GG.*viii*): though this is, for us, the crucial definition, Frege does not emphasize it in any way. Rather, he merely states it informally, by confining himself to stipulate that, for whatever object a,

'*a* belongs to the sequence of *Anzahlen* beginning with 0' (i.e. '$0\mathfrak{R}_{S^{-1}}a$', in my notation) means the same as '*a* is a finite *Anzahl*'.

Based on these definitions, Frege also proves a theorem corresponding to (SUCC_{GL}), which can be rendered as follows[15]:

$$\forall\psi, \chi[(\mathbf{N}x \wedge \mathcal{E}\psi = \overset{\iota}{\smile}\mathbf{f_I} \wedge \mathcal{E}\chi = \mathbf{f_I}) \Rightarrow x\mathfrak{R}_{\chi}\#\mathcal{E}[z : z\mathfrak{R}_{\psi}x]], \qquad (\text{SUCC}_{\text{GG}})$$

where also 'χ' ranges, as 'ψ', on first-level two-argument functions. Still, its proof and that of some corollaries is, again, followed by a long section devoted to *Endloss*, so as to suggest, like in *Grundlagen*, that his aim was less that of defining natural numbers, by singling them out among *Anzahlen*, than studying *Anzahlen* as such.

Despite the parallelism among these definitions and those of *Grundlagen*, the former do not identify natural numbers with the same objects as the latter. Whereas in *Grundlagen* natural numbers are identified with extensions of second-level concepts, in *Grundgesetze* they are identified with extensions of first-level concepts. Possibly, Frege was taking extensions of the former concepts to reduce to extensions of the latter. But, even if it were so, this would not depend on the definitions he offers, but on further informal stipulations, on which the very nature of natural numbers (and, more generally, *Anzahlen*) would, then, depend.

But, what is more relevant is that the definitions of *Grundgesetze* are even less apt to make natural numbers logical objects than those of *Grundlagen*.

On the one hand, similar considerations as those made above (pp. 16–19) about the definitions offered in *Grundlagen* also apply, *mutatis mutandis*, to the definitions of *Grundgesetze*. Indeed, according to these last definitions and BLV, 0 identifies with

$$\mathcal{E}[x : \mathfrak{T}_x \approx [z : z \neq z]],$$

while each positive natural numbers identifies with one of the following extensions

$$\mathcal{E}\left[x : \mathfrak{T}_x \approx \left[z : \bigvee_{i=0}^{i=n} z = i\right]\right] (n = 0, 1, \ldots)$$

and, in agreement with BLV and Frege's universalist perspective, these extensions are different in nature according to whether some non-logical facts obtain, unless it is circularly admitted that, among all first-level one-argument functions φ such that $\mathfrak{C}_{\varphi}$ is equinumerous to $[z : z \neq z]$

or $\left[x : \bigvee_{i=0}^{i=n} z = \mathfrak{i}\right]$, the only relevant ones for this matter are logical ones. For example, the nature of number 14, namely the extension of $\left[x : \mathfrak{T}_x \approx \left[z : \bigvee_{i=0}^{i=13} z = i\right]\right]$, cannot but depend, in Frege's views, on how many objects actually fall under the concept of eight-thousanders on Earth, or, at least, on which objects exists.

On the other hand, when one goes from the definitions of *Grundlagen* to those of *Grundgesetze* new problems arise for the identification of natural numbers with logical objects.

Consider 0. As we have just seen, it identifies with the *Anzahl* of $[z : z \neq z]$, namely the extension of the concept of being the value-range of a first-level one-argument function that takes the value $\top$ for no argument. Suppose there were plain objects, and that a be one of them. No object would, then, fall under $\mathfrak{T}_a$ so that 0 would be the extension of a (first-level) concept under which falls. Suppose there were not plain objects. Then, for any object, there would be a first-level one-argument function whose value-range is this very object, so that 0 would be the extension of a (first-level) concept under which fall all and only the value-ranges of a first-level one-argument function φ such that $\mathfrak{C}_\varphi$ is empty. It follows that the nature of 0 depends on whether there are or there are not plain objects. Provided that nothing within Frege's logical system can decide this matter, this makes it clear that this system cannot decide this nature, which makes, in turn, quite odd to take 0 to be a logical object because of the way it is defined within this system.

The supposition that there are no plain objects would not improve the situation. To see this, consider the following (first-level one-argument) function:

$$\mathcal{Q}(x) = \begin{cases} \text{The square constructed on } x & \text{if } x \text{ is a segment} \\ \mathcal{E}[z : z \neq z] & \text{otherwise} \end{cases}$$

Frege could not have taken it as a logical function. Still, unless he had come to identify $\mathcal{E}[z : z \neq z]$ with $\top$ (in opposition to what he says in §I.10), he should have admitted that $\mathfrak{C}_{\mathcal{Q}} \approx [z : z \neq z]$, which makes its value-range $\mathcal{E}\mathcal{Q}$, whatever it might be, fall under the concept whose extension is 0, by so suggesting that neither this concepts not this extension are logical items.

To avoid these two last difficulties, Frege might have defined 0 as the extension of the concept EXTENSION OF A CONCEPT EQUINUMEROUS WITH $[z : z \neq z]$, namely

$$\mathcal{E}[x : \exists X[x = \mathcal{E}X \wedge X \approx [z : z \neq z]],$$

under which, in agreement with BLV, only $\mathcal{E}[z : z \neq z]$ falls. This would have made it more plausible to take 0 as a logical object. Why did he not? The answer cannot be that his system involves no straightforward device to restrict quantification on second-order monadic variables to concepts. Since nothing would have forbidden him to invent such a device, or to conceive his system in a slight different way. Possibly the right answer is that he considered important to widen his concern from mere concepts to functions in general. Still, as worthy as this stance might be considered to be, it remains that nobody primarily aiming at identifying natural numbers with logical objects would have preferred the gain of generality it involves to the possibility of defining 0 as such an object.

But this is still not all. After 0, consider 1, and remember Frege's "stipulation" about value-ranges and truth-values of §I.10. According to BLV, this stipulation is equivalent to state that:

$$\top = \mathcal{E}[x : x = \top] \quad \text{and} \quad \bot = \mathcal{E}[x : x = \bot]$$

Compared with definitions (GG.*i*) and (GG.*v*), namely with

$$1 =_{df} \mathcal{E}[x : \mathfrak{T}_x \approx [z : z = 0]],$$

this makes 1 the extension of a concept under which both $\top$ and $\bot$ fall. Still, if another stipulation were made, namely if $\top$ and $\bot$ were not stipulated to be two value-ranges of a first-level one-argument function, or were stipulated to be value-ranges of such a function that does not takes the value $\top$ for a single argument, this would not be so. Hence, according to Frege's definition, 1 is a different object according to whether some stipulations are made or rejected. This would be quite strange for a logical object.

All this seems to show that the definitions of *Grundgesetze* are quite far from making natural numbers logical objects. By making these numbers, and, more generally, *Anzahlen*, extensions of openly identified concepts, which are, as such, independent of senses and intuition, they rather seem to aim at making truths about *Anzahlen*, and/or knowledge of them a priori, to ensure they necessarily exist, and to submit them to HP, so as to allow proving the basic theorems of arithmetic. Hence, as those of the *Grundlagen*, also the definitions of *Grundgesetze* suggest assigning to Frege a quasi-structuralist conception of arithmetic.

This does not mean that these last definitions would have not allowed Frege to argue that arithmetic is a part of logic. Since, if it is admitted that he endorsed such a quasi-structuralist conception, it should also be admitted that he might have considered that arithmetic is a part of logic, even if natural numbers are not logical objects. Indeed, he might have argued that this is so just because these numbers can be defined and arithmetical theorems proved within a system of logic. The informal definitions of *Grundlagen* could have hardly allowed him to argue this way. But the formal ones of *Grundgesetze* make this line of argument possible, provided, however, that a convincing argument for the logicality of BLV were offered.

7 A Last Remark

All this having being said, let me conclude with a counterfactual observation. It is easy to see that, once his system established, Frege could have easily defined natural numbers within it, in a quite different way, apt to make much more plausible taking them as logical objects. He could have defined 0 merely as the extension of the concept $[x : x \neq x]$, then adopted an appropriate definition of the extension of the successor relation, crucially but not greatly different from (GG.*iv*), and finally identified again the natural numbers with the objects that bear with 0 the weak ancestral of this relation, namely, with $\#\mathcal{E}[x : x \neq x]$, $\#\mathcal{E}[x : x = 0]$, $\#\mathcal{E}[x : x = 0 \vee x = 1]$, etc. Surely, the inability of his system to decide what value-ranges of first-level functions actually are would have made it impossible to say what these numbers would have been merely based on considerations internal to this system. Still, none of the previous considerations or other analogous to them would have made it hard to maintain that they are logical objects, provided that the system itself had been taken as logical.

So, why hid he not follow this route? A quite obvious answer is available: defining natural numbers this way would have made natural numbers neither submit to HP, nor be numbers of concepts, and would have so broken their crucial link with the more general family of *Anzahlen*.

This answer is as telling as obvious. Since it confirms that making natural numbers logical objects was not Frege's primary aim: submitting them to HP, and, so, making them numbers of concepts was much more important to him.

Notes

1. Cf. (Benis-Sinaceur et al. 2015, Chapter 1).
2. Here and throughout the paper, I use the terms 'arithmetic' and 'real analysis' to specifically refer, for short, to the theories of natural and real numbers, respectively.
3. Frege pervasively uses two German terms that might be translated in English with 'number': 'Zahl' and 'Anzahl'. The former is more generic, the latter more specific, and it is especially used to denote those numbers which provide an answer to how-many questions. This makes this latter term be often translated with 'cardinal number' (as in Frege GGAER). There are, however, several reasons—which, for sake of brevity, I cannot discuss, here—that make me think that this translation is improper (though not importantly misleading). This is why I prefer leaving this term untranslated. What is especially important to retain, here, about *Anzahlen*, is that Frege takes them to include natural numbers (and this is possibly his main point about the latter numbers). This makes all the claims about the former made in this § straightforwardly extend to the latter.
4. My quotes from Frege's works are substantially taken from to the current English translations of these works. Still, I will locally modifying them, if need be, in order the be more faithful to the original.
5. Remember that for Frege, a "truth" is analytic if (and only if) it can be proved from logical laws and explicit definitions alone (Frege 1884, §3).
6. This is what HP *prima facie* asserts in *Grundlagen* (Frege 1884, §73). To see that this is also what HP asserts in *Grundgesetze* is instead necessary to compare the (formal) statement of this principle (Frege 1893–1903, §§I.65, th. 32, and I.69, th. 49) with the informal explanations given earlier (ibidem, §I.40).
7. As odd as this phrase might appear, it is just Frege's own phrase translated literally (ibidem, §76).
8. A natural option would be to adopt a consistent variant of the formal system of *Grundgesetze*, powerful enough to allow us to offer an appropriate definition of natural numbers within it. To the best of my knowledge, among several available such variants, the one which is closest to the original system is the system PE, recently suggested by F. Ferreira (Forthcoming).
9. I have argued for that in (Benis-Sinaceur et al. 2015, Chapter 2).
10. Cf. footnote (8), above.
11. What I have in mind by admitting this possible pluralization of interpretations is Frege's acknowledgment that BLV is not enough for identifying value-ranges (cf. §4, above).

12. Namely to rephrase Frege's function $\xi \frown \zeta$, defined in §I.34 and, then, pervasively used in the whole work.
13. Cf. p. 9, above. Notice that if 'Δ' denoted, as for Frege, an object whatsoever, '$\mathfrak{T}_\Delta$' would denote, in my notation, the same concept Frege denotes with '$\_\_\xi \frown \Delta$', so that this claim coincides with that made by Frege at the end of §I.40.
14. Apart from Frege's passing from speaking of the "natural sequence of numbers" to speaking of the "sequence of *Anzahlen*", it is also noteworthy that the new relation is the inverse of the one considered in *Grundlagen*. Remark also that Frege conceives value-ranges of first-level two-arguments functions (including extensions of first-level binary relations) as value-ranges of appropriate first-level one-argument functions associated to these functions. So no extension of BLV is required for definitions (GG.*iv*) and related to be stated.
15. This is firstly stated, but not proved in §I.46. Its proof takes up all the first half of part II of Frege's treatise, and only ends in §I.118, where it is stated anew as theorem 155.

References

Benis-Sinaceur, H., M. Panza, and G. Sandu. 2015. *Functions and Generality of Logic. Reflections on Frege's and Dedekind's Logicism*. Cham, Heidelberg, New York, Dordrecht and London: Springer.

Boolos, G. 1998. *Logic, Logic and Logic*. With introductions and afterword by J.P. Burgess, ed. R. Jeffrey. Cambridge, MA and London: Harvard University Press.

Boolos, G., and R.G. Heck Jr. Die Grundlagen der Arithmetik, §§ 82–3. In M. Schirn 1998, 407–428. Also in Boolos, 1998, 315–338.

Carnap, R. 1929. *Abriss der Logistik […]*. Wien: J. Springer.

Carnap, R. 1931. Die logizistische Grundlegung der Mathematik. *Erkenntnis* 2: 91–105.

Carnap, R. PMBP. The Logicist Foundation of Mathematics. In *Philosophy of Mathematics*, ed. P. Benacerraf and H. Putnam, 31–41. Englewood Cliffs: Prentice-Hall, 1964. English translation of Carnap 1931.

Ferreira, F. (Forthcoming). Zigzag, and Fregean Arithmetic. Presented at the Workshop *Ontological Commitment in Mathematics. In Memoriam of Aldo Antonelli*, Paris, IHPST, December 14–15, 2015.

Fraenkel, A. 1928. *Einleitung in die Mengenlehre*. Berlin Heidelberg: Springer Verlag. Third edition reworked and largely extended.

Frege, G. 1884. *Die Grundlagen der Arithmetik*. Breslau: W. Köbner.

Frege, G. 1893–1903. *Die Grundgesetze der Arithmetick*, vol. I–II. Jena: H. Pohle.

Frege, G. GLAA. *The Foundations of Arithmetic*, trans J.L. Austin. Oxford: Blackwell, 1953.
Frege, G. GGAER. *Basic Laws of Arithmetic*, trans. and ed. P.A. Ebert and M. Rossberg, with C. Wright. Oxford: Oxford University Press, 2013.
Grattan-Guinness, I. 2000. *The Search for Mathematical Roots. 1870–1940. Logic, Set Theory and the Foundation of Mathematics from Cantor Through Russell to Gödel*. Princeton, NJ and Oxford: Princeton University Press.
Heying, A. 1931. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis* 2: 106–115.
Kant, I. WL. Wiener Logik. In *Gesammelte Schriften*, vol. XXIV.2, 785–940. Berlin: Walter de Gruyter, 1966.
Kant, I. LLY. *Lectures on Logic*, trans. and ed. J.M. Young. Cambridge: Cambridge University Press, 1992.
Panza, M., and A. Sereni. Forthcoming. Frege's Constraint and the Nature of Frege's logicism. To appear on *Revue of Symbolic Logic*.
Schirn, M. 1998. *Philosophy of Mathematics Today*. Oxford: Clarendon Press.
von Neumann, J. 1931. Die formalistische Grundlegung der Mathematik. *Erkenntnis* 2: 116–121.

Logic as Science

Robert May

1 The Flaw in Frege's Logic

Logical Realism and Anti-realism

It is a common reading of the genesis of Frege's conception of logic that there is a continuity, a line of refinement that runs from *Begriffsschrift* through *Grundlagen* to its finishing point in *Grundgesetze*. On this view, Frege's logical system was essentially in place in *Begriffsschrift*, later developments being tweaks and adjustments of the system introduced mainly for utility in developing the proofs of *Grundgesetze*.[1]

There is no doubt much truth to this perspective, but its acclaim has not been universal, and there are those who dissent that it gives an accurate read of the development of Frege's view of logic. A chief dissenter was Wittgenstein. In the *Tractatus*, he decries what he sees as a radical, and not for the better, shift in Frege's fundamental conception of logic, from logical anti-realism in *Begriffsschrift* to logical realism in *Grundgesetze*. This critical point is captured by Wittgenstein when he says "that there are no such things as 'logical objects' or 'logical constants' (in the sense of Frege and Russell)" (*Tr.* 5.4) and strikingly that

R. May (✉)
University of California, Davis, CA, USA
e-mail: RCMAY@ucdavis.edu

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_6

"My fundamental thought is that the 'logical constants' do not represent" (*Tr.* 4.0312). This stands in sharp contrast to Frege's treatment in his latter work, where these terms refer to concepts, and as such represent the truth-functions.[2] Wittgenstein's rejection of Frege's innovations in *Grundgesetze*, in favor of (refined versions of) prior treatments of *Begriffsschrift*, is thorough. Not only does he reject that truth-functions are concepts, but he also notably rejects that identity is objectual, that there are logical objects, and that sentences stand for objects (truth-values), rather than for complexes (propositions). In turn, Wittgenstein endorses that logic, as in *Begriffsschrift,* can generate a notion of ordering that is sufficient to induce the ordering of the natural numbers.[3] Overall, the *Tractatus* can be read fruitfully as a reflection on Frege's conception of logic, affirming what Wittgenstein sees as the logical anti-realism of *Begriffsschrift*, while rejecting the central realism of *Grundgesetze,* at the core of which is the notion that each properly formed expression of the conceptual notation has a reference in the conceptual hierarchy, inclusive of the logical terms.

If we are however envisaging the goal of the *Tractatus* as providing, largely via the picture theory, a philosophical foundation for the logic of *Begriffsschrift,* we would be somewhat off the mark. This would be because Wittgenstein's embrace of Frege's logic eschews one of its most notable features, which is common to the systems of *Begriffsschrift* and *Grundgesetze*. As Frege conceives of a logical system, it consists of a set of axiomatic propositions from which other propositions are derived as theorems. This directly implies that certain propositions are logically prior to others, conceptually more fundamental, standing as lawful statements which, in specifying the core content expressed by the system are in no need of derivation themselves. This Wittgenstein roundly rejects. On his view of logic, all propositions are on equal footing; none are conceptually prior to any others: "All propositions of logic are of equal rank; there are not some which are essentially primitive and others deduced from these" (*Tr.* 6.127). On Wittgenstein's view, to the extent that propositions are inferentially ordered, it is a reflection of an internal relation between their logical forms; it is not that the propositions of logic are derivationally ordered, starting with basic propositions, by external rules of inference. Accordingly, "Laws of inference, which — as in Frege and Russell — are to justify the conclusions, are senseless and would be superfluous" (*Tr.* 5.132).[4] Setting some logical propositions prior to others is not something that can be justified within logic.

If it is a mistake as Wittgenstein contends to think of logic as a system of axioms, which are laws of logic—that is, to think of logic in the format we think of a science—then this portends a profound problem for Frege's program. Frege's realism is intended to ground a scientific conception of logic, in the context of which the reduction of arithmetic is justified by providing an objectual foundation for the truths of arithmetic. Logic, on Frege's view, can have numbers as a subject-matter, whose laws can be derived within the logic from the basic logical laws. Wittgenstein for his part rejects that logic is in the business of stating laws of any subject-matter: Frege's realism sets a goal—or at least allows for a goal—for logic that is way too lofty, that logic itself could be a science. There is no dispute that logic can *serve* science. Rather, what is up for grabs is whether logic *itself* is a science. Can we ask about logic the same sort of foundational questions that are posed for physics or psychology? Wittgenstein's answer is a clear no. The realist stance is setting a standard that logic by its nature cannot meet.

The implication of this negative answer should not be missed: Frege is *unjustified* in taking the Basic Laws of logic as foundational. They are not, and cannot be, scientific laws, and as such cannot ground scientific truths or our knowledge of those truths. Accepting this claim would cast high suspicion on the conceptual coherence of Frege's logic, and more broadly on his logicist program. While the particular logical system that Frege presents in *Grundgesetze* is inconsistent, the critique here is more profound: It is that Frege's theory of logic, with its scientific aspirations, is conceptually flawed, and deeply so.

So what are we to make of this? One stance would be to see the dispute as a kind of faultless philosophical disagreement. Frege and Wittgenstein are each just relying on their respective realist/anti-realist base presuppositions, and so talking past one another: Those with a taste for realism will hardly be swayed by anti-realism, and vice versa, for obvious reasons. My sentiments reside here, for reasons that will emerge. However, others have seen something more telling. To them, Wittgenstein's critique invites an argument that Frege's logical theory cannot meet its own *internal* standards for what constitutes a scientific law. If logical laws are scientific laws—like the laws of physics, they are laws of what *is* true[5]—then whatever standards justify the laws of physics must be the same sort of standards as justify the laws of logic. The argument is that Frege cannot make good on this requirement: While scientific laws are *judgements*, comparable justification is not forthcoming for

the laws of logic. Frege may place the judgement stroke at the head of the logical laws, but this is illegitimate, an act of pure hubris.

The importance of evaluating this argument is plain: If Frege is mistaken that logic can provide foundational notions and principles, then his logical realist program is at risk, along with his logicist vision of mathematics. But are these accusations just? I think not. To the contrary, on Frege's view of logic *as* science, the logical laws *are* justified in the manner of proper scientific laws. That is to say that within Frege's general conception of science the demand for judgement *is* met. The remainder of this paper is devoted to showing just why this is so.

The Science of Logic

In his paper "Formal Theories of Arithmetic", published just after *Grundlagen*, Frege makes this remark[6]:

> First, no sharp boundary can be drawn between logic and arithmetic. Considered from a scientific point of view, both together form a unified science.

In sounding a familiar logicist theme, this remark is by no means unique in Frege's writings. Nevertheless, it is notable in the way that it casts this theme, as scientific unification, that no sharp divide can be drawn between the *science* of logic and the *science* of arithmetic. One might deny the status of science to one or the other of logic or arithmetic, but this was not Frege's way of thinking. To him, both are fully worthy of their scientific credentials; logicism, Frege insisted, was a fully scientific thesis. But not only this. Logicism is also a reductionist thesis; the unification of arithmetic and logic is effected by the reduction of arithmetic to logic. Accordingly, the science of logic is the prior science. Even more strongly on Frege's view, it is the most basic, and most general, of all the sciences.

The question of the unified science that occupies Frege is thus just the question of what it is for logic to be a science. Fully answering this question will require us to pry rather deeply into Frege's conception of logic (we will come to this below), but at the outset there is one thing we can say about it; viz. what it does *not* mean for logic to be a science: To say that logic is a science is not to say that it is part of psychology, *qua* the science of thinking, even if logic is a characterization of reasoning in

some very fundamental sense. To place logic in this context is to commit the sin of psychologism. Frege is firm on this: Logic is a science in its own right, although with a special relationship to the other sciences.

Now, if logic is an independent science, then there must be something that it is a science of: What are the things about which lawful properties are being asserted within the science? Here again there is something we can say. If logic and arithmetic are a unified science, then whatever is the subject-matter of arithmetic must also be the subject-matter of logic. Accordingly, the scientific subject-matter of logic are numbers, the subject-matter of arithmetic. This too is a familiar theme of Frege's logicism, but its familiarity does not relieve the contentiousness of the idea that logic has a subject-matter at all, no less than that part of that subject-matter are numbers.

We return to this concern. But first an elaborative remark is called for about one other implication of Frege's view already mentioned, that logic is the most basic of sciences. By saying this, Frege means to be understood as saying that logic is part of, embedded within, all other sciences; all other sciences have logic at their core. Because of this the subject-matter of logic (granted that it has one) will be part of the subject-matter of all sciences, and so too will be the logical laws that govern this subject-matter. In this way logic thought of scientifically gains a certain generality via its being an embedded part of any scientific theory whatsoever. This too is not without its contentiousness, and will require further elaboration, to which we will return as well.[7]

Putting these characteristics together, the following observation emerges: On Frege's view of logic, it is not just that the Basic Laws of logic specify the lawful properties of any conceptual universe (*i.e.* any universe that conforms to the conceptual hierarchy), but that any such universe is also an arithmetic universe, *by definition*. By this definition, numbers are objects of conception (value-ranges of concepts) governed by the Basic Laws; accordingly, the Basic Laws of logic taken as scientific axioms are ultimately the axioms of arithmetic.[8] The justification for this claim resides in that truths of arithmetic are derivable from the Basic Laws; that is, that they are *judgements* that can be derived.

"The aim of proof," Frege tells us in *Grundlagen*, "is, in fact, not merely to place the truth of a proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another" (2). This dependence is made transparent by logical derivation, formalized as steps from judgements to judgements, each step in the sequence

legitimized by a rule of inference. Backtracking through this sequence will reveal "the ultimate ground upon which rests the justifications for holding [a proposition] to be true" (*Gl*, §3). So too:

> the fundamental propositions of arithmetic should be proved, if in any way possible, with the utmost rigour; for only if every gap in the chain of deductions is eliminated with the greatest care can we say with certainty upon what primitive truths the proof depends. (*Gl*, §4)

Ex hypothesi, for arithmetic truths this will lead back to the Basic Laws of logic; that is, arithmetic truths are justified as judgements because they logically rely on *axiomatic* judgements, judgements that are themselves *underived*.[9]

It is at this point that we engage the issue. If judgements are ultimately justified in virtue of their derivations, then it would appear that those propositions that are underived must be unjustified, or if justified, justified in some other manner than by derivation. Plausibly scientific laws are of the second sort, being empirically justified. But the Basic Laws of logic are not like this; they are not even more broadly synthetic judgements, by Frege's own admission. But if so, then short of some other non-derivational mode of justification, we would need to recognize that Frege's view of logic as science has been called into question, and in a very substantial way, for it is based on the following claim: Frege's view of logic as a science implies that the Basic Laws of Logic are *unjustified*; more strongly, they are *unjustifiable*. And as such, they do not provide a foundation for arithmetic within the context of Frege's reductionist program.

The Demand for Judgement

If logic is to be a science, then it must have a subject-matter; just like any other science it must be about something, the things of which its laws are the laws of. But what might this be? One thought is that logic has a specific subject-matter all its own, like physics or chemistry. This is Frege's thought: numbers are the subject-matter of logic. This is a strong form of logicism, that logic has numbers as its particular and peculiar subject-mattter.

At face value, the claim of strong logicism seems arbitrary and unjustified: Why should numbers be the peculiar subject-matter of logic?

Why not points and lines, or matter or the cosmos, for that matter? Why should numbers be singled out from all the rest? Doesn't logic hold in all of these domains equally? Frege himself seemed to think so, at least if he is sincere when he remarks that the "basic propositions" of logic "must extend to everything that can be thought" (Frege 1984a, p. 112). Taking this as a central dogma, it appears to be a mistake to think that logic has a specific subject-matter. Rather, it has a *general* subject-matter: Logic's applicability is not delimited to a particular domain; it applies to all. The subject-matter of logic is unrestricted, and it does not discriminate in what falls under its laws. As such, these laws are absolutely general; there is nothing that does not fall within their scope. They would be universal laws, and logic would be the science of all things.

This conception of logic has been attributed to Frege; a particularly clear brief is found in Conant (1991). In describing Frege's conception of logic, he tells us that: "The laws of logic hold for anything, any sort of subject-matter whatsoever",[10] and amplifies this remark with the comment[11]:

> What the laws of logic do continue to lack, on Frege's view, is a subject-matter that is specialized in any way; their subject-matter is simply: everything.

If the subject-matter of logic is absolutely everything, then obviously numbers cannot be its unique subject-matter. Arithmetic falls under logic because everything falls under logic. But this rather takes the stuffing out of logicism: Yes, numbers are a subject-matter of logic, but then so too is everything else, and this makes for a very weak form of logicism.[12]

Now weak and strong logicism may differ over whether the subject-matter of logic is general or specific, but they do agree on one thing, that logic *has* subject-matter. To this extent, it is not in dispute that logic is a science. But there is another way of thinking of the topic neutrality of logic other than that it is universal that sheds this commitment: It is not because logic applies to everything, but because it applies to nothing. The inspiration is from Wittgenstein: "The logical propositions describe the scaffolding of the world, or rather they present it. They 'treat' of nothing" (*Tr.* 6.124). Logic is maximally general because it has *no* subject-matter; it is, in itself, about nothing. It is equally applicable to reasoning about any subject-matter whatsoever, being the framing on which any particular reasoning hangs. As such, logic aids and abets the

generation of knowledge, but logic alone cannot generate knowledge. Its truths, as Wittgenstein says, are always tautologies, and "therefore say nothing"; there can be no logical knowledge per se. But then arithmetic knowledge cannot be a type of logical knowledge, if there is no genuine logical knowledge in the first place.

From the perspective of this Wittgensteinian stance towards logic, what becomes highly suspect is the idea that logic is a science. That we are warranted in this skepticism, as Conant argues, is shown once it is observed that under Frege's assumptions, the axioms of logic cannot be judged as truths. If logic is a science, then it must generate truths that can be judged as such, and hence known. These truths include the laws of the science, its foundational axioms. If logic is a science, then its axioms too must be judged truths; the Basic Laws of logic must be justified truths just as much as the basic axioms of any science. But, Conant maintains, because of Frege's methods of making judgements, the Basic Laws of logic are *unjudgeable*, since we cannot for them entertain whether they are true *or* false. They cannot meet what Conant calls the "demand for judgement".

Conant presents his argument as follows[13]:

> That is, there isn't any sense to be made of the idea of someone (even God!) entertaining the falsity of a basic logical law. And this, in turn, would mean that Frege's account of judgment fails to leave room for anything which could count as judging a basic law of logic to be true. The demand for judgment, in the case of the axioms of *Begriffsschrift*, would turn out to be unintelligible. Yet Frege's account of logic as the maximally general science requires that we be able to judge the axioms of his system to be true. If we are to conceive of the laws of logic as differing from those of the other sciences only in their order of generality, then they must be able to serve as possible candidates for judgment. So Frege's view that the basic laws of logic possess positive content does not afford any basis for their inability to face the demand for judgment.

As Conant sees the matter, the problem is that we cannot pose the judgemental question—Are they true or not?—of the Basic Laws. This is tantamount to saying that the Basic Laws of logic do *not* have judgeable content; *because* they do not have judgeable content, they are not the sorts of thoughts of which we can make a judgement. That would be "unintelligible". They are not *proper thoughts*, the sort of thoughts

subject to judgement. But if so, they are no better than *mock* thoughts, no better than fictions, and can give at best an illusion of knowledge; we surely would not think of mock thoughts as generational of knowledge. *Ergo*, logic cannot be a science.

Frege distinguishes two senses of what it means to justify a truth: (i) a truth justified by its *acknowledgement*, and (ii) a truth justified by its *dependencies*.[14] Truths justified in these ways are *judgements*.[15] By (ii), obtaining a judgement depends on the relations of a judgement to other judgements, as this is revealed by its proof. But not so by (i). This requires a cognitive process, of recognition of truth, and this is to have a psychological explanation.[16] With this distinction, we can move to the following question: What is it for a truth to be *logically* justified? By (i), it is to be logically acknowledged; by (ii), it is to depend on—to be derived from—logical laws. In turn, this leads us to the further question: In which of these senses are the Basic Laws of logic justified? By (ii), they are *un*justified, obviously so, since they are axioms and hence not derived, and so not dependent on any other distinct truths (setting aside the irrelevant trivial derivation by reiteration). So the question devolves on whether the Basic Laws are justified by (i), and here Conant's answer is a decisive no. There can be in principle no recognition of the truth of the Basic Laws of logic, and so they cannot be granted the status of judgements.

To be clear here, to give the negative answer is to say that Frege's conception of logic is rotten to the core. It is not just that he was mistaken that one of the Basic Laws was a truth (*i.e.* Basic Law V), he was mistaken that it was even possible to establish that they are truths in the first place. It is not just that the Basic Laws are unjustified as judgements, it is that they are unjustifiable. Claiming that logic is a science places too heavy an epistemic burden on the Basic Laws of logic. This, according to Conant, is what Wittgenstein recognized about Frege's logic, that Frege made a serious blunder in thinking that logic has epistemic foundations sufficient for the generation of substantive scientific knowledge: Logic is *not* a science.[17]

The upshot of this line of thinking is that there is a flaw in Frege's logic; its locus resides in the linking of the doctrine of logic as science to the strong logicist thesis. This is incompatible with Frege's adherence to the universality of logic, and leads to the conclusion that the Basic Laws of logic cannot meet the "demand for judgement", and so fail to be scientific laws. But is this dire reading of Frege's project justified? I think

not; explaining why will depend on two clarifications. The first will be of Frege's conception of logic; the second will be of his conception of judgement. In the context of these clarifications the affirmative answer to the question posed above will take root: The Basic Laws of logic can be judged, and so justified in a way suitable for the foundation of a scientific inquiry, with all of its attendant epistemic force. So we have our task, to show that the Basic Laws of logic *do* meet the "demand for judgement", as Frege understands this notion.

2 Frege's Conception of Logic

The "Pure" Logic of Conceptual Content

There is no doubt that Frege sets a certain epistemic claim at the conceptual center of his thinking. We can put it as follows: To come to have knowledge of, and reason about, an object, we must "see" it as the argument of a particular sort of function, of a concept.[18] By doing so, it becomes possible to establish truths about that object, and it is by apprehension of such truths that we come to have knowledge of that object and reason about it at a standard appropriate to scientific inquiry. Clearly, it is this thought that is embodied in Frege's Context Principle.

Now Frege took this principle to be completely general. *Any* domain of objects, no matter what they might be, analyzed conceptually this way—as a structure of function and argument—is, for Frege, potentially a logical content, that is, a content of judgements. In his earliest writings, in *Begriffsschrift,* Frege had an apt term for content understood this way: He called it *conceptual content.*

In *Begriffsschrift,* Frege announced that conceptual content is "that which is of sole importance" for logic; accordingly, his goal was to develop a "pure" logic that can apply to any content whatsoever, so long as it can be construed as conceptual, *i.e.* as analyzable as a structure of function and argument. This logic, Frege tells us, will "transcend all particulars" and will be "independent of the particular state of things".[19] Accordingly, logic takes on a certain universality befitting its ontological neutrality: The laws of logic hold good for *any* conceptual content, and so include the conceptual content of scientific domains, the important case to Frege's mind.

Alongside his view about conceptual content, Frege held another view that is key to understanding his epistemology. It is this: *We have no direct*

cognitive access to content. In saying this, we do not mean to attribute to Frege an absurd view, that we do not have perceptual (or perhaps intuitive) awareness of content. Rather, it is to say that we have no comparable awareness of the *logical structure* of content. Hence, we have no direct access to the *truths* of content. But without some sort of access we can have neither knowledge of, nor can we reason about, the objects of our awareness, perceptual or otherwise.

What is needed is thus some way that we can conceive of the logical structure of content, some sort of conceptual system via which we can grasp logical structure. We require, and this is Frege's key insight, a notation by which conceptual content can be *represented*, and so made accessible to us. We need a "conceptual notation"; we need a Begriffsschrift that represents content as a structure of function and argument.[20] Via the Begriffsschrift, as Frege is given to saying, content can be *viewed* as *carved* into a structure of function and argument, so that by grasping the Begriffsschrift, we can grasp the logical structure of content. Here we arrive at a distinctively Fregean theme, that our access to content is through language, given that language is an expressive, representative medium.

The Begriffsschrift as it is being described is as found on its initial presentation in 1879 in Frege's book of that title. As presented there, the Begriffsschrift is a representational system that represents solely and strictly by *display*. That is, the structure of the conceptual notation *transparently shows* the logical structure of content, and so represents in a strong sense logical form. The Begriffsschrift makes formally apparent, as something that can be "viewed", the logical structure of content; it does so by displaying it as compositions of functions and arguments. It is in terms of these representations that inference can be characterized as rigorous, "gap-free" proof, as logical derivations. The logical rules of derivation thus do not operate directly on contents, but rather on their representations, mapping judgements, *qua* representations of true contents or facts, onto other judgements.[21]

What then is represented by:

$$f(a)?$$

Frege would say that what is symbolized by "f" is a function and what is symbolized by "a" is an argument.[22] That they compose is indicated by the horizontal stroke that "ties the symbols which follow it into a whole"[23]:

$$— f(a).$$

Frege calls this the judgeable content. That this judgeable content is affirmed is indicated by the addition of the vertical stroke:

$$\vdash f(a).$$

This represents an affirmed truth, a judgement.

$$\vdash\!\!\!\top f(a)$$

is a denial; *qua* judgement, it affirms the truth of the negation of the proposition. It is on this conception of the Begriffsschrift that Frege seeks to fulfill what he takes to be goal of purely logical analysis: To show that truths of conceptual content can be derived from basic lawful properties of those contents.

Notably on this way of thinking, the Begriffsschrift is decidedly *not* compositional, not in the semantic sense of this notion. It is only at the very end does anything in the way of semantic notions enter the picture, after the function and arguments have been tied together into a whole by the horizontal stroke. The vertical judgement stroke, Frege says, "relates to this whole"; whether a judgement is an affirmed or denied truth is *not* something that is a function of the semantic values of its parts. The Begriffsschrift *represents* the truths of conceptual content, and in a logical system formalized in the Begriffsschrift it can be shown how they are related by logical laws. This goal only requires a notion of formal composition which governs proper substitutions in proofs: The goal is not to show how those truths are determined by the meanings of the components of that truth. If it were, there would be the need for semantic clauses to particulars; there would need to be reference clauses. But then logic would no longer be transcendent in Frege's intended sense. Such clauses would be at odds with the ontological neutrality of logic. The absence of semantic clauses in the presentation of *Begriffsschrift* is notable; indeed, notable is the absence of the notion of *bedeutung* itself, as it becomes familiar in Frege's later work.[24]

The point then amounts to this. While logic is universal in applying to any content (subject-matter) conceptually construed, it is not referentially anchored in any. The pure logic of conceptual content assumes no particular ontology, and so there is also no particular semantics; this is a reflex of ontological neutrality. If we ask what the terms of the conceptual notation refer to, the answer is going to be: nothing in particular.[25]

The logical perspective of *Begriffsschrift, qua* representational system, is if anything formal and syntactic, and not semantic. The introduction of a rich semantic framework, in which reference to particulars is possible, is going to have to wait. But not for long.[26]

Descriptive and Ontological Logicism

Against this background of the pure logic of conceptual content, Frege poses a question in *Begriffsschrift*: What can we know of content, only with the understanding of it as being conceptual? This is tantamount to asking what we can know "as a matter of logic". Frege's concerns are rather broader, however, and he wants an answer to an even more pointed question: What can we know just as a matter of logic that is of mathematical significance? Here is Frege's answer[27]:

> I had first to test how far one could get in arithmetic by means of logical deductions alone, supported only by the laws of thought, [i.e. of logic], which transcend all particulars. The procedure in this effort was this: I sought first to reduce the concept of ordering-in-a-sequence to the notion of logical ordering, in order to advance from here to the concept of number.

On Frege's view, to say that the mathematically significant notion of ordering is a logical notion is to say that this notion can be defined with respect to *any* conceptual content, and moreover that with respect to conceptual content, fundamental theorems can be proven about such orderings. Frege tasks himself these proofs in the third part of *Begriffsschrift*. It is this part of Frege's logicism that George Boolos labeled "sub-logicism".[28]

As Frege puts it in the remark above, it is necessary to resolve first the sub-logicist issue of pure logic—what aspects of arithmetic are reflections of general properties of conceptual content?—"in order to advance from here to the concept of number". The separation of tasks to which Frege alludes is not intended to reflect merely a rhetorical division of labor, between the tasks of sub-logicism and logicism proper, but rather something deeper, something that occasioned, Frege tells us, "a deep reaching development in my logical views" (Frege 1893/1903, p. X). The cause of this development is that the logic of *Begriffsschrift* is unable to be the logical system in which logicism can be carried through. This logic is

universal; if in any sense it is about everything, then it is about anything, and hence not peculiarly about numbers. Accordingly, the results developed in this context are applicable to arithmetic, but are independent of it. But to make the advance to which Frege alludes, it is necessary to show something stronger, that logic is peculiarly about numbers. But not only this. It is to show that that logic is about numbers in the manner of being about a scientific subject-matter. This is key, if Frege is to make good on his claim that logic and arithmetic form a unified science.

In *Grundlagen,* Frege turns to the advance to the concept of number. The core idea presented there is that while it is concepts that have numbers, numbers themselves are particulars, "self-standing objects". But what sort of objects are these? Clearly in the context of logicism as a reductionist project, we need some way of specifying what they are in a strictly logical fashion. In *Grundlagen*, Frege canvasses two possibilities.

The first is just to say that these objects are the sorts of things that can have numerical properties; that is, that they are whatever it is that falls under numerical concepts, where these concepts themselves can be logically characterized. So long as there is a criteria in place for individuating these objects, itself given in logical terms, it might be thought that there is a clear advantage to this view, as it does not load up logic with any strong metaphysical claims. But this form of *descriptive* logicism, Frege rejects as too weak: It is the view presented in *Grundlagen* §§62–65 that entertains Hume's Principle as an implicit definition, and which is rejected in §§66–67 on the grounds that as a definition Hume's Principle is insufficiently general.[29]

In its place, Frege opts for a stronger view of *ontological* logicism. It is not just that numbers are objects that can be logically described, it is rather that numbers are a particular sort of logical object: *extensions* of concepts. Among these independent, self-standing logical objects, certain ones can be isolated and defined as cardinal numbers. Which ones? This is specified by an *explicit* definition (as opposed to the implicit definition above).[30] This definition, in terms of equivalence classes of equinumerous concepts, Frege first presents in *Grundlagen* §68:

(#) The number of $F =_{df}$ extension of the concept *equinumerous with F.*

Continuing through §86, Frege first shows that the *definiens* of (#) is given in purely logical terms, equinumerosity being defined in terms of one-to-one correspondence, and then on the basis of (#), sketches proofs

that there is an infinite sequence of cardinal numbers; that is, that (#) affords the characterization of the class of natural numbers.[31]

Frege is clear, however, that he has only indicated how to obtain this result[32]:

> … it can still always be doubted whether they are deducible solely from purely logical laws, or whether some other type of premise is not involved at some point in their proof without our noticing it. This misgiving will not be completely allayed even by the indications I have given of the proof of some of the propositions; it can only be removed by producing a chain of deductions with no link missing, such that no step in it is taken which does not conform to some one of a small number of principles of inference recognized as purely logical.

Ten years hence, Frege's goal in *Grundgesetze* is to allay this doubt, formally presenting the result in the format of rigorous, gap-free proofs. He does so with a critical elaboration: The term "concept" is to mean a function whose values are always truth-values, and so the term "extension" is to mean the value-range—the pairings of arguments and values—of a concept. With this emendation, Frege modifies the definition so that what is defined is not the number of a concept *F*, but rather the number of its value-range. That too is a value-range, of the concept of being a value-range of a concept *equinumerous with F*. But in giving the definition this way, Frege does not abandon the idea of *Grundlagen* that numbers are objects, and that number-terms refer to these objects. Rather, if anything he has refined the idea that numbers are objects of conception (not objects of perception or intuition) by identifying them with the value-ranges of concepts.

As noted, the logic of *Begriffsschrift* is unsuited for logicism, since to the extent that it is a compositional system, it is only in a syntactic sense: Conceptual content is formal content. The innovation of *Grundgesetze* is the introduction of a compositional *semantics*, expressed as the conceptual hierarchy: Each term of the Begriffsschrift has reference in the conceptual hierarchy.[33] Notably, the horizontal stroke is to be understood not solely in virtue of its syntactic role, as in *Begriffsschrift*, but as a concept (mapping the True to the True and everything else to the False), and the other logical terms follow suit as truth-functions. For this construction to give a notion of concept, however, it must be founded at the 0-level with entities that themselves are not concepts. Frege's trick—and

what he took as his great insight—was to claim that value-ranges are the 0-level entities.[34] But note: The only substantial properties of value-ranges that interest Frege are those that contribute to the characterization of numbers, in particular, the properties in virtue of which they can be placed in a natural sequence. There are value-ranges that are not numbers in this sense, but Frege has no interest in deriving any particular truths about them in *Grundgesetze*.[35] Rather the goal there was to derive the Basic Laws of Arithmetic, *i.e.* the laws of numbers.[36]

Ontological logicism is obviously a much stronger claim than descriptive logicism, as it commits logic to embedding a definite metaphysical claim; as a reductive program, it makes logic out to be peculiarly about numbers by identifying numbers with certain specified logical objects. But what is Frege's justification for this stronger claim? It is this: It allows that logic can have a subject-matter in a manner no different than any other science, that numbers are the particulars of this scientific subject-matter, and that the Basic Laws of logic are the axioms of this subject-matter. It is this version of strong logicism that Frege adopts in *Grundgesetze*. But this does not capture all the nuance that Frege has in mind: Frege's idea is that all sciences are *applied* logics. Arithmetic is special only in being the first of these, a science so basic that it is a component of all other sciences. Seeing how this is is our next topic.

Applications of Logic

What then is Frege's conception of science? It is this: A science is an *application* of logic to a subject-matter.

On Frege's view, we can characterize any particular scientific application of logic by a specification of three things: (i) A *domain* of entities, given by a "source of knowledge"; (ii) *Axioms* that govern these entities; and (iii) *Terms* that refer to the entities *qua* subject-matter of the science. Logic sits at the core of any such application, placing the most fundamental constraint on content, that it be conceptual. This is the domain of pure logic, as specified by the Basic Laws of logic. To the Basic Laws are added, for any given application, a set of axioms that stipulate the basic lawful properties of the subject-mattter of the application, along with definitions that introduce terms for the concepts and objects of that subject-matter.[37] Frege's conception of science is accordingly of an axiomatic system with a subject-matter, where the deductive base, for any given science, comprises the Basic Laws, the axioms that govern the

subject-matter and the definitions, and where inferences are carried out in accordance with the rules initially given within the core logic. It is this last demand that ensures the degree of rigor that is required of scientific inquiry, and founds the justification of scientific claims by their dependency on the fundamental laws of the science. All of this is to be carried out in the proprietary notation—the Begriffsschrift—which guarantees the formal rigor of the science.

To give an example, geometry is an application of logic. It is logic having as its subject matter geometric entities, to wit, points and lines, whose basic properties are determined by the Euclidian axioms, the laws of the geometric entities that are given by intuition.[38] These axioms, along with appropriate definitions, added to the Basic Laws of logic give geometry as a scientific application of logic. This method for constructing applications, with pure logic at its core, is completely general throughout the sciences, inclusive of those whose subject-matter is given empirically; parallel remarks could be put into place for physics, chemistry, psychology, etc.

Frege's conception of science incorporates a sharp distinction between the logic of conceptual content and the applications of that logic; between pure logic and applied. Making this distinction, however, requires an important clarification in how the Begriffsschrift, the logical language, is to be understood as a representational system. Pure logic, recall, has no particular subject-matter, and this ontological neutrality is reflected in the Begriffsschrift having no semantics to speak of. But this is decidedly not the case for applied logics. They have particular subject-matters, and their propositions are to express truths of that subject-matter. In an applied logic, it simply will not do to say of "*f(a)*" just that *f* is a function and *a* is an argument; we must say more, that "*f*" stands for a particular concept *f*, that "*a*" stands for a particular object *a*, and that in virtue of standing for these things, their composition is a truth (or not). Applied logic demands that the Begriffsschrift be symbolic of subject-matter, of its particular concepts and objects. The Begriffsschrift must have the representational capacity not only for conceptual content, but also for subject-matter content.

What this means for the Begriffsschrift *qua* language is that its signs are to be complexes, part formal symbol standing for conceptual content, and part representation of the particular that is signified. Its sentences will be composed of signs in this sense, hence simultaneously representing both logical structure and objectual content. Frege's way of putting

this is to say that a sign expresses a sense, and that for every composition of signs into a sentence, there will be a corresponding composition of senses, what Frege calls a thought: In this regard, a sentence expresses a thought. While the formal structure of the sentence displays its logical structure, it is the thought expressed that determines what that sentence is about, and it is thoughts "for which the question of truth can arise at all" (Frege 1977c, p. 4). They are true or false as a function of the references of their component signs.

Throughout Frege's work, the role of the Begriffsschrift as a representational system remains a constant, with an invariant epistemic role of representing content in a cognitively accessible way.[39] What advances from the presentation in *Begriffsschrift* to that of *Grundgesetze* is the content represented; from conceptual content to conceptual content with a subject-matter. This is the advance from pure logic to applied logic. It is this advance, at the heart of Frege's conception of the place of logic in science, that presses the development of Frege's semantic conception: Applied logics require that the Begriffsschrift have a compositional semantics, even if pure logic does not.

Arithmetic as Applied Logic

The question now looms: How does arithmetic fit into this scheme, given Frege's view that arithmetic is the science of number? As a science, the thought is that it would be constructed as is any other science, with a specification of the axioms that govern its subject matter. For arithmetic, Frege's thesis is just this: That the axioms *are* the Basic Laws of logic. Arithmetic is thus the application of logic that results from the *null* addition of axioms to the Basic Laws. The thesis is that logic can be about numbers without any extensions or emendations of the Basic Laws.[40] As with any other science, logic applied in this way will have scientific content in virtue of what can be proven from the axioms about its subject-matter, that is, about numbers. Logicism, as now articulated, is the thesis that arithmetic is the *first* application of logic, with numbers as its subject matter, and in which the basic laws of arithmetic, the truths about number, are provable as logical theorems from the axioms of the theory.

At this point, Frege's commitment to ontological logicism becomes understandable, for it is this view of logicism that supports the reduction of arithmetic as scientific unification. The task for Frege is to characterize

within logic numbers as a scientific subject-matter, and it is by a definition of number that such a domain is carved out. For this task, descriptive logicism is insufficient; specifying numbers as *whatever* meets certain conditions is too weak. If it is left open ontologically exactly what numbers are, it cannot be determined in virtue of their definition whether or not they are the sorts of things that can be given as a subject-matter. We would in a strong sense fail to know just what are the things of which scientific claims are being made, and hence the truth of scientific propositions. Frege decries such a circumstance in no uncertain terms. It would be "really a scandal that science is still in the dark about the nature of number", without "any agreement about the meaning in arithmetic of the word "identical" and the sign of identity." The consequence would be that "Science … does not know what thought content is associated with its theorems or what subject matter it deals with: it is completely in the dark about its own nature. Is this not a scandal?"[41]

To avoid this, Frege advises, we must be explicit about what numbers are; this is what is accomplished by the explicit definition of number (#) by identifying certain value-ranges as numbers. These are logical objects; our grasp of logic is sufficient to know of their existence, and to be able to individuate them as objects of conception.[42] Numbers are given, as Frege would say, by the "logical source of knowledge", and, by Frege's lights, this is as sufficient to characterize a scientific subject-matter as are other sources of knowledge, intuition for geometry, or empirical for the non-mathematical natural sciences.[43] It is key to the dialectic of *Grundlagen* that neither the intuitive nor the empirical can support a coherent notion of number, and that this is only afforded on a logical basis, and this is taken for granted in *Grundgesetze*. While Frege does hold that scientific reduction is a virtue in itself, his argument is in fact stronger. Only by reducing arithmetic to logic can we locate entities that suffice to properly characterize numbers in a way that they are subject to scientific inquiry; *i.e.* that they can be characterized as a scientific subject-matter.

In the context of arithmetic *qua* application, as far as pure logic is concerned, what is represented is that "$2+3=5$" and "$3+4=7$" are both true, both are judgements, and that they have the same logical form. In the applied logic, this will be preserved. But beyond this it will also be represented that the first is a truth about the number 5, that it is the sum of 2 and 3, while the second is a truth about the number 7, that it is the sum of 3 and 4. These are facts of the subject-matter, of arithmetic. They are logical truths because the axioms of the applied logic

are logical laws; that is, they are *logical* truths in the same sense that we speak of *geometric* truths or *physical* truths. All to Frege are species of scientific truths.

On Frege's conception, the applicability of logic places it at the core of science, and makes it part of any science whatsoever; all sciences are built around logic. The immediate consequence is that scientific domains are always conceptual domains, and that what constitutes proper scientific thought—thought whose quest for certainty requires that it meet the highest standards of rigor—can be specified by the logical laws of inference. As we have just seen, arithmetic is no exception; logic is as much at its core as it is of any other science. Arithmetic, however, is exceptional in one way; it travels arm-in-arm with logic. Its laws occur in all other sciences, as do the laws of logic. Frege is clear this is no coincidence[44]:

> As a matter of fact, we can count just about everything that can be an object of thought … What is required is really no more than a certain sharpness of delimitation, a certain logical completeness. From this we may undoubtedly gather at least this much, that the basic propositions on which arithmetic is based cannot apply merely to a limited area whose peculiarities they express in the way in which the axioms of geometry express the peculiarities of what is spatial; rather, these basic propositions must extend to everything that can be thought. And surely we are justified in ascribing such extremely general propositions to logic.

The invitation to the unification of arithmetic and logic can be met on Frege's conception of science because logic can be an application to *itself*: Frege's logicism is the thesis that this application is arithmetic. A science is an applied logic, so it is *as* arithmetic that logic is a science, so to speak. This science, the unification of logic and arithmetic, has numbers as its subject-matter, and the Basic Laws of logic as its axioms. It is this science that is at the core of scientific inquiry, and it is by this embedding in science that the "demand of arithmetic that its numbers should be adapted for use in every application made of number, even although that application is not itself in the business of arithmetic"[45] can be met. Arithmetic, Frege is ultimately telling us, *is* logic, when logic is viewed in a scientific light: "Considered from a scientific point of view, both together constitute a unified science."[46]

The Universality of Logic

On Frege's view of logic and science, logic sits at the core of science; pure logic is at the core of all of its applications. At the core of science thus sits what is true of every conceptual content, that is, of every content that can be characterized as a structure of function and argument. The generalizations about conceptual content that are theorems of pure logic remain in force when moving to an applied logic; they remain in place when functions and arguments are interpreted in terms of the particular concepts and objects that constitute a scientific subject-matter.

Now against this conception, what does it mean to say that logic is universal? There are at least two senses that can be distilled. On the first, the laws of logic are about everything, in that anything—concept or object—is either a function or an argument, and so falls under the laws of conceptual content. In this regard, pure logic has a universal subject-matter, but in an ontologically neutral way; entities fall under the laws of conceptual content for no other reason than that they are concepts and objects. Nothing specific about them enters into the matter. On the second sense, logic is universal by being embedded in all domains of inquiry. Pure logic is the core of each and every applied logic; accordingly, the logical laws of conceptual content are maximally general because they occur in every science. The subject-matter of logic in the broad sense is the universal closure of all of the subject-matters of its applications; that generalization ranges over everything is a closure property over the totality of the applications of logic. In each of the applications however, generalization is restricted to the subject-matter of the application; its range is limited to the objects (0-level entities) and concepts (higher-level entities) of that application.

The importance of distinguishing these senses comes home to roost with arithmetic. The Basic Laws, *qua* the laws of conceptual content, are just as much part of arithmetic as of any other science; to the extent that their theorems are independent of scientific content, so too in the context of arithmetic. But by hypothesis, the Basic Laws are also the axioms of an application, and in this regard have a restricted subject-matter. It is numbers that fall under the Basic Laws of logic *qua* the axioms of arithmetic, just as it is points and lines that fall under the axioms of geometry, or elementary particles fall under the laws of physics. As scientific laws, the Basic Laws are no more universal than the laws of geometry or

physics, even if, given the reductionist tenet of logicism, they are carried along with logic itself into every scientific application.

For Frege, the central concern is not so much logic in itself, but rather logic as it makes a scientific contribution. That the laws of logic are laws of conceptual content is of importance because given their application, they impose that scientific domains are conceptual domains, and hence obey the standard of reasoning demanded by scientific inquiry. That the laws of logic are the axioms of arithmetic is of interest because from them the truths of number are derivable in this manner. It is in these ways that we can take logic substantively as a science, but if so, we still have an issue. If the laws of logic are to be understood as scientific laws of the same order as geometric laws or physical laws, then presumably they are to be justified in the same way that other scientific laws are justified. But we have in place a claim that the logical laws cannot be justified in this way; that they cannot meet the "demand for judgement". So, we move on; With the clarification of Frege's notion of logic as science in hand, we move on to his understanding of judgement and its role in epistemic justification.

3 Judging and Judgement

If there is a demand for judgement, then what is that is being demanded? What is being asked for? To provide an answer, we need to consider Frege's account of judgement, and how in particular we are to understand those propositions that Frege writes as follows:

$$\vdash p.$$

To be sure, Frege says different things at different points. In *Begriffsschrift*, this expresses a judgement as a matter of predication; the initial symbol headed by the vertical judgement stroke stands for the notion "is a fact", and is the "common predicate for all judgements". Frege comes to reject this way of thinking of judgement, however, and accepts rather the opposite. Thus, in Frege's considered view, which emerges at the time of *Grundgesetze* and continues through the remainder of his writings, judgement is not to be taken in this semantical way. In contrast to the other logical strokes—the horizontal content stroke, negation and the conditional—the vertical judgement stroke "does not serve, in conjunction with other signs, to designate an object" (Frege 1970a, p. 34, fn.). It is this view, on which the judgement stroke is

syncategorematic, and judgements are not a distinctive category of propositions, that frames what we will call Frege's theory of judgement.

We can gain an initial flavor of what Frege has mind by the following remark from *Grundgesetze* §5 by way of introducing the judgement stroke[47]:

> I distinguish the judgement from the thought in such a way that I understand by a judgement the acknowledgement of the truth of a thought.

It is this acknowledgement that is marked by appending the judgement stroke; the notation above thus represents that *p* is a thought recognized to be true. A judgement we can say is a thought annotated as judged.

There is looseness of usage in this statement that Frege is intent to clarify. He wants to distinguish a judgement, as described, a thought with an annotation, from *making a judgement*, the process via which the truth-value of a thought comes to be recognized. *Judging* in this sense can be thought of as annotating or appending the judgement stroke, and how Frege understands what this amounts to is clarified by the following remarks, which are typical in Frege's later writings[48]:

> The act of judging is a psychical process, and as such it needs a judging subject as its owner. ("Negation")

> Both grasping a thought and making a judgement are acts of a knowing subject, and are to be assigned to psychology. But both acts involve something that does not belong to psychology, namely the thought. (Notes for Darmstaedter 1919)

> When we inwardly recognize that a thought is true, we are making a judgement; when we communicate this recognition, we are making an assertion. ("Logic" 1897)

> When someone comes to know something it is by his recognizing a thought to be true. For that he has first to grasp the thought. Yet I do not count the grasping of the thought as knowledge, but only the recognition of its truth, the judgement proper. ("Sources of Knowledge of Mathematics and Mathematical Natural Science")

There are two main take-aways from these remarks. First is that making a judgement is an agentive, psychological act; cognitive events that are located in space and time. Without an event of judging, there can be no judgement. Once a judgement is made it may persist, and may be

transmitted in its manifested form as an assertion.[49] Judging on Frege's lights is a cognitive act, matched with assertion, a speech act.[50]

The second matter of note is that making a judgement is a knowledge generating activity (and assertion is knowledge conveying); it is the generative source of knowledge. By recognizing via an act of judging that a thought is true, one thereby comes to know that the thought is true: A judgement is a known truth. Judging, however, does not make thoughts be true, nor does it create or alter thoughts[51]:

> Judging (or recognizing as true) is certainly an inner mental process; but that something is true is independent of the recognizing subject; it is objective. (Letter to Jourdain)

> Our act of judgement can in no way alter the make-up of a thought. We can only acknowledge what is there. A true thought cannot be affected by our act of judgement. ("Negation")

> A traveller who crosses a mountain range does not thereby make the mountain range; no more does the judging subject make a thought by acknowledging its truth. ("Negation")

Epistemically, acts of judging are reflective acts on what there is.

The ensuing question is then what are the objects of judging acts; to what do the apply? Answer: They apply to *judgeable contents*. Initially, in *Begriffsschrift* Frege spoke of the judgeable content as that which is either affirmed or denied by an act of judgement. But in *Grundgesetze*, he revises the notion, as part of "a deep-reaching development in my logical views"[52]:

> Previously, I distinguished two components in that whose external form is a declarative sentence: 1) acknowledgement of truth, 2) the content, which is acknowledged as true. The content I called judgeable content. This now splits for me into what I call thought and what I call truth-value. This is a consequence of the distinction between the sense and the reference of a sign. In this instance, the thought is the sense of a proposition and the truth-value is its reference. In addition, there is the acknowledgment that the truth-value is the True. For I distinguish two truth-values: the True and the False.

The judgeable content is now to be understood as having two parts; they are pairings of thoughts and truth-values. These pairs are the relata of a

semantic relation, that of sense and reference. To make a judgement is to recognize that the relata do, as a matter of fact, stand in that relation.[53]

For Frege, our initial cognitive encounter with a thought is via our grasp of it, where by grasp of a thought we have direct and immediate awareness of that which is thinkable: "To think is to grasp a thought", Frege tells us.[54] But if grasp acquaints us with a thought, it does not thereby also acquaint us with its truth: "Truth is not part of a thought" in Frege's estimation, so that "We grasp a thought without at the same time recognizing it as true — without making a judgement."[55] Because of this, although grasp may be sufficient for apprehending what can be known, it is in general not sufficient for that knowledge: "the mere thought alone yields no knowledge, but only the thought together with its reference, i.e. its truth value."[56] It is by judging that these are brought together, the further psychological act by which an agent acknowledges or recognizes the truth of a thought.

By the grasp of a thought, an agent is placed in a position to judge that thought, and in principle come to have knowledge. What this ability amounts to is that an agent may entertain the question, for the thought that p, whether it has the judgeable content (p, the True) or (p, the False). "We grasp the content of a truth", Frege says, "before we recognize it as true, but we grasp not only this; we grasp the opposite as well."[57] At this point a decision has to be made: Which is it? Which one of p and *not-p* is true, and which one is false[58]:

> When we are concerned with the truth of a thought, we waver between opposite thoughts, and with the same act we recognize one as true and the other as false.

Judging, then, is to figure out which it is. It is to make a choice[59]:

> To each thought there corresponds an opposite, so that rejecting one of them coincides with accepting the other. To make a judgement is to make a choice between opposite thoughts. Accepting one of them and rejecting the other is one act.

Thus a thought, Frege tells us, is "that to which the question 'Is it True?' is in principle applicable."[60] To judge that p is to answer the question "Is p true". If the answer is affirmative, then we have the judgement

that *p*; if the answer is negative, then we have the judgement that *not-p*. To ascertain the proper answer is to gain knowledge.

We can now state Frege's theory of judgement as follows: For (*T*, ?), T a thought, the agent's task is to solve for ?, given *T,* where ? is either the True or the False. Thus, to judge that *T* is to answer the question "Is *T* true?".[61] If the answer is affirmative, then we have the judgement that *T*; if the answer is negative, then we have the judgement that *not-T*. To ascertain the proper answer is to gain knowledge.

Put a little differently, but equivalently, we can say that for any thought *T*, there are two judgeable contents:

(*T*, T): Represented as: — *T*

(*T*, F): Represented as: ⊤ *T*,

with the question of judgement being: To which can the vertical judgement stroke be appended? There are two possibilities:

⊢ *T*: The judgement that *T*

⊢ *T*: The judgement that *not-T*

While in the first case *T* is a true thought and in the second a false thought, it would be improper to label the first as a true judgement and the second a false judgement. A judgement is *always* a truth: That *T* is true *or* that *not-T* is true. And just as there are no true or false judgements, there are no true or false judgings. Truth-values are not the sort of notions that apply to acts of judging; rather, they are applicable to the thoughts so judged.

Apprehension of a thought—*grasping*—places an agent in a position to undertake a cogitative act—*judging*—and come to know that a thought is true; the thought itself is the objective basis for a subjective, knowledge generating, cognitive act. But since grasping a thought is prior to judging, grasping underdetermines knowledge: Awareness of the first parameter of a judgeable content—the thought—is not in general sufficient to determine the value of the second parameter—the truth-value. Judging is figuring this out. It is a discovery of an existing state, that a thought is paired with a truth-value *as a matter of fact*. Once discovered, this can be acknowledged, and the form of the thought can

be annotated with the judgement stroke. Accordingly, an agent who has judged the thought that p, knows that p is a way of thinking of the True. On this picture, then, there are two basic epistemic attitudes, thinking and knowing, and to say that an agent thinks or knows a thought would be to say that the agent has grasped or judged that thought. Thinking is prior to knowledge, and necessary for it; only if a thought has been grasped can it be judged.[62]

The acknowledgement afforded by judging is, in Mark Textor's apt phrasing, ontic; it is to recognize one of two objects, either the True or the False.[63] Thus, given that judgeable contents are sense/reference relata, to make a judgement is to recognize the truth-value to which a thought refers. But if judging is ontic in this sense, the *grounds* for making a judgement are not; rather, they are conceptual. To recognize that a thought refers to the True it must be recognized that there is predication; at the first level that objects falls under concepts (and at higher levels that concepts fall within concepts): To recognize that $P(a)$ refers to the True, it must be recognized that a falls under P, where information needed for the identification of the concept and the object is provided by the senses that compose the judgeable thought. Recognizing the truth of a thought is thus a form of conceptual understanding; to make a judgement is to gain conceptual knowledge, the semantic reflection of which is the ontic relation of thought to truth-value, as an instance of the relation of sense and reference. Commonly to figure out whether there is predication, we need some form of sensory/perceptual experience, be it in scientifically controlled environments or in everyday life. As Frege observes, "A sense impression is not in itself a judgement, but becomes important in that it is the occasion for our making a judgement" (Frege 1979i, p. 267). But not always. Sometimes our grasp of the concept P is itself sufficient to know that for any a, $P(a)$ refers to the True, and hence that $P(a)$ is a judgement. It is important to bear this special case in mind; its importance will shortly become clear.

We now have sufficient grasp of Frege's theory of judgement to move on to the question at hand: Can the logical laws be justified as judgements, and so meet the demand for judgement? This, as we can now see, amounts to asking whether we can make these judgements. What would be judging a Basic Law amount to in the context of Frege's theory of judgement? That is our next topic.

4 Judging the Basic Laws

If we think of modeling the theory of judgement, then making a judgement can be understood as a mapping from judgeable contents to judgements. Thought of this way, judgements are the result of successful acts of judgement, and judging so conceived is a process by which knowledge is generated by coming to know that a thought is true. This is to be contrasted with a different mapping which is distinctive of logic: Derivational mappings in the Begriffsschrift from judgements to judgements. While the outputs are the same, the inputs are not. The latter mapping is formal, governed by rules of inference that, if rigorous and reliable, are proof supporting. Judging in way of contrast is a psychological activity, and as such cannot meet the rigor and gap-free reliability standards required for proof. It is prone to error, and although it may perhaps be formalized as a way of theoretically modeling psychological activity, this is not essential, as it is for logic.[64] To conflate deriving with judging is to commit the sin of psychologism, to confuse laws of logic with laws of thinking, and Frege spends considerable time warning us off this mistake.[65]

That derivational transitions from judgements (not judgeable contents) to judgements are inferential steps, Frege makes clear[66]:

> Logic is concerned only with those grounds of judgement which are truths. To make a judgement because we are cognizant of other truths as providing a justification for it is known as *inferring*. There are laws governing this kind of justification, and to set up these laws of valid inference is the goal of logic.

Logical derivation makes us aware of judgements by revealing their containment in other judgements, "as plants are contained in their seeds" as Frege prosaically puts it.[67] What can be derived need not be obvious, however: "What we will be able to infer cannot be inspected in advance", but will only be revealed by the sound construction of proofs that synthesize interactively from core judgements; a derived proposition "consequently is not contained in any one of them alone, yet does follow purely logically from all of them together."[68]

How this goes Frege makes quite clear. A logical system (including applied systems) starts with a core set of judgements—axioms and definitions that can be premisses for any derivation whatsoever—and that set

is then extended in derivations via rules of inference. A simple example: The rule of detachment—the core inference rule of Frege's system[69]—derives from judgements of the form:

$$\vdash \begin{array}{l} B \\ A \end{array}$$

and:

$$\vdash A$$

the judgement:

$$\vdash B.$$

The effect of detachment is to add the conclusion to a body of judgements just in case both of the premisses are already in the body of judgements.

"The conclusions we draw" from logical derivations, Frege tells us, "extend our knowledge". The knowledge obtained this way may be novel: There may be thoughts that we can come to know only by revealing through proof their immanence in prior judgements: If we know that *A* implies *B*, and we know that *A*, then our knowledge can be extended to knowing that *B,* given that if we have the judgement that if *A* then *B*, and if we have the judgement that *A,* then the judgement that *B* can be derived. By this logical accretion of knowledge, core judgements can be amplified into whole bodies of knowledge. But bear in mind that this knowledge is of a special sort. Although "it not uncommonly happens that we first discover the content of a proposition, and only later give the rigorous proof of it" (Frege 1884, p. 3), even if via derivation is the only way in which a proposition could be known, from logic we only obtain knowledge relative to the dependence of that proposition on others. It is this sort of second-order knowledge that is provided by logic. The epistemic novelty provided by a proof "is not the content of the proposition, but how its proof is conducted, on what foundations it rests" (Frege 1893/1903, p. XIII). Derivational knowledge, so to speak, is distinct from (first-order) judgemental knowledge.

What logic provides is an interconnected web of dependencies among judgements, and hence among knowledge. Proof is the method for the construction of this nexus, and it "afford[s] us insight into the dependence of truths upon one another."[70] For Frege, the Begriffsschrift is a logical system that accomplishes this axiomatically: Ultimately,

the dependencies can be traced back to a class of core judgements—axioms and definitions—from which they are sourced.[71] But a question looms. Frege's logical goal is not simply to establish methods of deriving sentences from sentences—a *calculus ratiocinator*—but rather to develop a system for showing the interconnections of knowledge—a *lingua characteristica*—by deriving judgements from judgements. The applicability to scientific knowledge is based on the relations of derived judgements to the core judgements that in part define the science. The question then is the following: What is the source of the core judgements that found any particular nexus? How are they made? Given that logical rules of inference map judgements into judgements, it follows that all the derived judgements are also made, once the core judgements are made. So the issue devolves upon how this occurs. Obviously, it can not be by being derived: The point of being an axiom is precisely that it is not derived within the system. Thus, the core judgements must be made in some other way.

In this passage, Frege appears to allow two other ways that this might be accomplished[72]:

> Now the grounds which justify the recognition of a truth often reside in other truths which have already been recognized. But if there are any truths recognized by us at all, this cannot be the only form that justification takes. There must be judgements whose justification rests on something else, if they stand in need of justification at all.

Here Frege acknowledges that while there are judgements that are inferentially justified, there are others that are justified non-inferentially, that is, not by being logically derived. For these, which for the reasons just outlined are inclusive of the core logical judgements, there are two possibilities. Either they are self-justifying as judgements, or they are not; if not, they "rest on something else." But if not, there is only one other way: The core judgements must be made in the sense of judging. That is, that they are in fact judgements can *only* be established by showing the transition from their judgeable content to the judgement proper. I will return to the first alternative momentarily, in deference to the second. The reason is that this latter alternative is the one Frege explicitly adopts.

How so? The core judgements of logic are of two sorts. The Basic Laws, by which we mean the six Basic Laws of *Grundgesetze*, which are the logical axioms, and the definitions, of which there are seventeen in

Grundgesetze.[73] Of the two sorts of core judgements, we may dispose of definitions. So long as the *definiens* are expressible within the proprietary vocabulary of the science, in the case at hand, logical terms, we can help ourselves to these for free: It is a consistent feature of Frege's logic that for any definition whatsoever, there is always a corresponding judgement which may occur as a premiss in any proof.[74] In Frege's notation, this is a conversion from:

$$\Vdash A = B,$$

where A is the definiens and B the definiendum, to the judgement:

$$\vdash A = B.$$

Since definitions are stipulative, this is all there is to say about the source of the latter judgements.[75]

Axioms are a different matter. They are substantive truths, not stipulations, and so the question arises of how they are judged and hence how we come to know these truths. Frege's answer we have already given: We must judge them, just like any other substantive truths. The grounds for doing so, however, are circumscribed by the science of which they are the axioms, most narrowly for logic. The axioms of logic—the judgements Frege singles out as the Basic Laws—are to be logically judged, that is, without reliance on anything non-logical.

This is how Frege proceeds, explicitly in *Grundgesetze,* §18[76]:

> According to §12
>
> $$\vdash\!\!\left|\begin{array}{l}\Gamma\\ \Delta\\ \Gamma\end{array}\right.$$
>
> would be the False only if Γ and Δ were the True while Γ was not the True. This is impossible; accordingly
>
> $$\vdash\!\!\left|\begin{array}{l}a\\ b\\ a\end{array}\right.$$

The displayed judgement is Frege's Basic Law I. What Frege shows here is the transition from judgeable content to judgement—note the lack of the judgement stroke on the first formula—and he specifies the grounds for making that judgement. Given the specification of the conditional as material—this is the reference to §12—there can be no instantiation of the first formula that is false: The truth-value of every

judgeable content of the specified form must be the True. Accordingly, the generalization—Basic Law I—must be a judgement.[77] This is the way Frege proceeds for each of the other Basic Laws; in §18, for Basic Laws IV and VI, and in §20 for the remaining Basic Laws, II, III and V. In each case, there is a transition from judgeable content to judgement, and although there is variance in the bases for judging each Basic Law, the commonality is that they are each judged as following from the stipulations governing the primitive logical functions, inclusive of the truth-functions, identity, generalization and description.[78]

At this juncture, we pause to revisit Conant's remark cited above, "that Frege's account of judgment fails to leave room for anything which could count as judging a basic law of logic to be true. The demand for judgment, in the case of the axioms of *Begriffsschrift*, would turn out to be unintelligible."[79] It is hard to see how to reconcile this with what Frege says quite intelligibly in *Grundgesetze*. Conant is right that we cannot "entertain the falsity of a basic logical law"—that just follows from the Basic Laws being judgements. But we certainly can entertain the falsity of its judgeable content—that the thought is paired with the False—and this is exactly as we must, and as Frege does, in making the judgement.

Of course, there can be no (non-trivial) judging of the Basic Laws if making a judgement is to mean giving a formal proof in the Begriffsschrift. But this commits the mistake of conflating proving with judging. Judging is not a derivational process; it is certainly not a synonym for logical proof. Not surprisingly, Frege declares that the Basic Laws of Logic "themselves neither need nor admit of proof"—obviously not, since they are axioms.[80] But this is not to say, or even imply, that they neither need nor admit of being *judged*. To the contrary. On Frege's view, "it is part of the concept of axiom that it can be recognized independently of other truths",[81] that is: judged. Since the Basic Laws are a species of axioms, the logical axioms, this goes for them as well. They are not proven, but they are judged. They meet the demand for judgement.[82]

What then has Frege accomplished? What he has *not* done is establish that Basic Law I (and the other Basic Laws) *is* a basic law of logic. At most his business is to establish only that it satisfies a necessary condition for being a Basic Law of logic, that it is a truth, and moreover, given that the grounds for judging are logical, that it is a logical truth. There is a different question: Is there some property of thoughts—perhaps

self-evidence or unprovability—that makes them eligible to be Basic Laws? But this is not Frege's question. He is not seeking a sufficient condition.[83]

5 Unification, Not Reduction

At the outset of our discussion, the goal was set to show not only what we just observed—that the laws of logic are justified in being judged—but also that this justification is in the manner of the justification of scientific laws. This is necessary if logic is to be a science, and given the tenets of logicism, to ground arithmetic as a scientific inquiry. This claim then raises the last question of this paper, whether the judgements of the logical laws have been made in a scientific manner, or rather in some way that is distinct, and in particular, distinctively logical.

As described by Frege, judging the logical laws does have its definite characteristic: Upon reflective evaluation of the functions on which logic is founded, and the ways in which those functions can be composed with arguments, it will be apparent that the Basic Laws are logical judgements. This is what will be "obvious as must properly be required of a logical law" (*not* that they are Basic Laws of logic).[84] That they are judgements will be self-evident in the sense of evident from the logic itself, as an intrinsic feature of the structure of the logical system.[85] But not all axioms are judged like this: It may not be equally apparent from the structure of an applied logic whether its (non-logical) axioms are judgements. This may be possible to establish only on extrinsic grounds, empirically or, as Frege would have it, by intuition (in the case of geometry).

Now, it may seem that this is just a way of saying that the Basic Laws are analytic of the logical concepts, and are judged a priori, in contrast to the axioms of applied logics, which are judged a posteriori: Basic Laws of logic are a priori because they are lawful manifestations of primitive concepts of the logical system, and that they are so judged will be apparent. In this light, Frege's logicism is most naturally seen as a project of epistemic reduction. The goal is to show that arithmetic truths are secured epistemically by being logical theorems, and so a priori, with the implication that they are epistemically distinct from truths of other sciences.

This evaluation of Frege's goal misses the mark, however. Frege's goal is not to show how the logical laws are different from other scientific laws, but rather to show what they have in common in the way they are

judged. For Frege, judging a thought is *always* a matter of conceptual analysis. If it is successful, the result is conceptual knowledge, viz. that an object falls under a concept. In the case of scientific laws, the concepts will be the fundamental concepts of the theory. In §18 and the following sections of *Grundgesetze,* Frege analyzes the Basic Laws in this way. In these cases, the concepts are logical concepts: For instance, it follows that Basic Law I is a judgement because any objects *a* and *b* fall under the concept[86]:

The issue is not so much the epistemic character of the logical laws, but that they are judged in the manner of scientific judgements.[87] For certain, the information deployed in making those judgements will be relativized to the science whose laws are subject to judging. For logic this will be non-empirical, since on Frege's conception logic is prior to other sciences, including physics (so that the grounds for judging will be independent of space and time). But this is ultimately irrelevant. Rather, what is relevant is whether the Basic Laws can be judged on the basis of conceptual analysis, of objects falling under the foundational concepts of the science. Frege's argument as presented in *Grundgesetze* is that the Basic Laws of logic are properly axiomatic in this sense. To the extent that logicism is a reductionist program, it is based on the claim that the laws of arithmetic cannot themselves be judged directly; that neither intuition nor empirical concepts can provide the grounds to judge. (This is a central rhetorical theme of *Grundlagen*.) The derivation of the arithmetic laws from the logical laws however will be sufficient to show that they are judgements, given that the logical laws are judgements.

As a matter of textual accuracy, Frege speaks in *Grundlagen* of the logical laws, and hence the arithmetic truths derived from them, as analytic; cf. §§3*ff*. But two things are important to bear in mind in evaluating these sections. One is that while Frege uses the terms analytic and synthetic, he does not do so in the traditional Kantian sense, in that the category of a judgement is something that can be discerned from inspection of its logical form. Frege's rejection of this thesis is fundamental; it is central to Frege's logical insights that the notion of logical form that supports rigorous proof will *not* manifest this distinction. Rather, Frege appropriates the distinction in order to categorize types of proofs, so as to isolate those that depend on logical laws and definitions and nothing else.

Second, and importantly, at the point of *Grundgesetze* and in subsequent writings, Frege ceases referring to arithmetic truths as analytic. Rather he designates them as logical truths. This is not a mere shift of terminology, but rather reflects a sharpening of Frege's core thinking about logic to emphasize their status as a species of scientific truth. Thus, the term "logical truth" is to be understood in a manner comparable to "physical truth" or "chemical truth". Just as the laws of physics are justified on physical grounds, or those of chemistry on chemical grounds, the laws of logic are justified on logical grounds. They are each species of scientific truths. In this regard, Frege's concern with logic as science is a the level of the genus, not the species.

Given Frege's focus on the justification of scientific propositions, it is no wonder that his concern with the epistemic category of judgements, whether they are a priori or a posteriori, withers away, and is ultimately of lesser if not little importance in his later work. What is of importance is the psychological act of judging, and to this the epistemic categories are inapplicable.[88] Logicism is primarily a project of scientific unification, not epistemic reduction, and central to the thesis is that the Basic Laws of logic are axioms, and judged accordingly. The judgements derived from the axioms will constitute scientific knowledge, individuated by their subject matter. Arithmetic truths are a type of scientific truth, and judgement of them will be scientific knowledge of numbers. Frege's concerns are epistemological to the extent that they are reflective of an underlying scientific project, of a concern with the acquisition and judgement of scientific knowledge. On Frege's considered view, logical knowledge, in the form of arithmetic knowledge, is an instance of scientific knowledge, its acquisition and justification governed by the principles that govern the acquisition and justification of scientific knowledge in general.

6 A Concluding Remark

To conclude, we return to the opening of this paper and ask what we should think of Wittgenstein's take on Frege? Did Frege fundamentally change his view about logic between *Begriffsschrift* and *Grundgesetze*? Did he give up logical anti-realism to embrace realism, and by doing so embrace a flawed conception of logic, a flaw that undermined his conception of logic as science? At the least, we have taken the sting out of this accusation. If the litmus test is to meet the demand for judgement,

then Frege passes. But if this is beyond dispute, nevertheless we need to recognize that it is in virtue of presupposing that the basic logical concepts are givens that are ultimately justified by the overall role they play in the logical construction. But pushing matters back like this would hardly satisfy the critic; it is precisely the existence of these entities as referential values that the anti-realist rejects. Here we have reached the philosophical faultless disagreement that was alluded to earlier. A realist could no more give up the existence of the logical concepts than an anti-realist could accept them. We have entered a philosophical *cul-de-sac*.

Frege we should grant was a logical realist from the very beginning, and remained so for the duration. But saying this is not meant to belie that there are not real and substantial differences in Frege's two presentations of logic. If anything, these reflect Frege's channeling of Kant's distinction between formal and transcendental logic: Simplistically, Frege's update is that the latter is the former with a compositional semantics. It is expressed in the distinction between core and applied logic, where core logic is characterized in the formal mode, applied logic in the material. The most idiosyncratic aspects of *Begriffsschrift*—the treatments of identity and generalization—arise from Frege attempting to hew to the formal while eschewing the material: Frege's initial insight is that in this context substantive theorems can be derived applicable to the development of arithmetic in logic. But Frege clearly understands at the outset that this is not the whole story; it cannot be given the limitation of the formal that logic is not anchored to any particularity. Rather, it is the applications of logic that provide anchoring, via the innovation of the semantic framework developed in *Grundgesetze*. It is in this framework that Frege's conception of logic as science takes full form.

Acknowledgements It is an honor to have this paper appear in a volume for Eva Picardi. Eva was a profound scholar, an exceptional person and a great friend. Her influence runs through the fabric of this paper. Versions of this paper were presented at the Frege Workshop, IUSS Pavia, May 2015, the France-Berkeley Workshop, Université Paris I, December 2015 and at the Next Generation Research Promotion Project, Global Research Centre for Logic and Sensibility, Keio University, Tokyo, November 2016. Thanks are due to those audiences, and to Sanford Shieh for very helpful discussion.

Notes

1. Chief among these innovations being the introduction of value-ranges and the definition of converse of a function.
2. Most clearly, the distinction at hand can be garnered by Frege's differential treatment of the horizontal. In *Begriffsschrift*, it plays a purely syntactic role, binding together into a unity the function and argument terms that follow it. In *Grundgesetze*, it stands for a truth-function that maps the True to the True and everything else to the False. More on this later.
3. This is accomplished by the introduction of the notion of an iterative operation; cf. *Tr* 6.02. See Floyd (2001) for an illuminating discussion.
4. To the extent that rules of inference are countenanced, they are as props to allow the internal inferential relations to be displayed. They do not, however, in any way determine what can be inferred from what.
5. See Frege's remarks in the Forward of *Grundgesetze*, p. xiv *ff.* where Frege contrasts this view of logical law to that of the psychological logician, for whom logical laws are laws of being *taken* to be true.
6. Frege (1984a, p. 113).
7. Bear in mind that by speaking of logic being embedded in or part of a science, it is meant only that they are *principles* of every science, and so neutral on whether they are axioms or just presupposed propositions.
8. Roughly, Frege distinguishes three types of scientific subject-matter that can be given for analysis: Those containing objects of perception (empirical sciences), objects of intuition (geometry) and objects of conception (arithmetic). See Frege (1979i).
9. Obviously, setting aside trivial reiterative proofs from themselves, since any proposition can be proven this way.
10. Conant (1991, p. 138).
11. Conant (1991, p. 139).
12. While Conant is particularly forthright in his attribution of this universalist perspective to Frege, he is by no means alone in subscribing to this understanding of Frege's conception of logic; cf. Goldfarb (2001).
13. Conant (1991, p. 140).
14. Note that in *Grundlagen* §3, Frege reserves the term *justification* for the second sense. The first sense Frege calls *making a judgement*, or judging. Care in usage will emerge as we proceed.
15. To be clear, for Frege judgements are known *truths*; accordingly there can be no false judgements: There can be no judgement of false thoughts, although there can be of their negations. Frege makes this point explicitly in "Negation" (Frege 1977b).
16. Throughout his later writings, Frege is clear that he sees making a judgement as acknowledging or recognizing that a thought is true, and in

numerous places he states that this is an agentive, psychological act. We return to this in Sect. 3.

17. There are responses on Frege's behalf. One is that the Basic Laws are self-evident truths, and so we just *know* that they are true, Another is that the Basic Laws are primitive truths, their epistemic content given by pre-logical elucidations. See Jeshion (2001) for the former, Weiner (1980) for the latter. Reasons to doubt that either are Frege's view are located in Sect. 4.
18. The claim extends: In order to come have knowledge of and reason about concepts we must see them as arguments of higher-level concepts.
19. Frege (1972, p. 104).
20. For clarity, I adopt the following convention: "Begriffsschrift" in roman letters is used for the logical language; in italics for Frege's book.
21. These remarks naturally harken to Wittgenstein's distinction between saying and showing, although in the *Tractatus* §5.53*ff,* Wittgenstein imposes an even stronger transparency condition than Frege. For Wittgenstein, every difference in the logical symbolism reflects a difference in semantic value; consequently, "$a=b$" is necessarily false (or not well-formed in the language, and hence non-sensical), as the symbolic difference itself implies that "a" and "b" have different references. Frege holds no such thing. His notation is intended to represent logical *form*: *f(x,y)* represents the same logical form regardless of whether x and y are assigned the same or different values.
22. This simplifies. What Frege says is that "we call the part of the expression that shows itself invariant a function and the replaceable part its argument" (Frege 1972, p. 127). So it could be that "a" is the function and "f" is the argument. See Heck and May (2013) for discussion of how this relates to Frege's treatment of generalization in *Begriffsschrift*.
23. Frege (1972, p. 112).
24. Because of the way that semantic notions enter in *Begriffsschrift*, it is only in terms of affirmation and denial that Frege can characterize the truth-functions, that is, in terms of judgements. What he says is that:

$$\vdash\!\!\!\begin{array}{l} A \\ B \end{array}$$

"stands for the judgement" that A is denied and B is affirmed "does not occur", but that the other three combinations of affirmation and denial do. That is, we have the displayed judgement just in case we either have the denial of B—the judgement of the negation of B—or the affirmation of A—the judgement of A. See *Begriffsschrift* §5. To the extent that this anticipates the truth-tables, it is only in *Grundgesetze* that the material conditional is definitively characterized as a truth-function; see

Grundgesetze §12. This clarification required, among other innovations, a compositional semantics that is not to be found in *Begriffsschrift*; cf. Heck and May (2018).

25. Hence an implicit sort of generality is built into the Begriffsschrift: It is that in *f(a)*, "*f*" stands for *any* function, and "*a*" stands for *any* argument. Frege explicitly notes this, and indicates it notationally by the use of Roman miniscule letters. Note that to indicate a different function or a different argument, we would have no recourse but to use different letters. But a difference in lettering in itself does not guarantee a difference in function or argument (although identity of lettering does guarantee identity). That the typographical distinction symbolizes the same function or argument is assertable in the presentation of *Begriffsschrift*. This is to assert that the distinct symbols have the same conceptual content; that is, that they are logically non-distinct. For more, see Frege's discussion of identity of content in *Begriffsschrift* §8. Also, see fn. 21.
26. Until *Grundgesetze* in a logical setting, although it already begins to emerge in the early 1880s in Frege's unpublished paper comparing his logic to Boole (Frege 1979b). This development of Frege's thought is discussed in Heck and May (2013). Note that one way of thinking about *Begriffsschrift* is that Frege only presents propositional logic, not predicate logic. This immediately raises the question of the treatment of generalization. In Heck and May (2013), it is shown that Frege's treatment of generalization in *Begriffsschrift* is also non-semantic; his fundamental insight here is just the representation of scope as a syntactic difference. The semantics of generalization waits for the development of the conceptual hierarchy, in which (first-order) quantifiers are second-level concepts.
27. Frege (1972, p. 104).
28. See Boolos (1998b).
29. This is on the basis of the so-called Caesar argument: It does not follow from the definition whether Julius Caesar is or is not a number. See May and Wehmeier (2018) for recent discussion of Hume's Principle and Frege's conception of definition.
30. For discussion, see Wright (2016), who defends implicit definition in the context of neologicsm; for a response, see May and Wehmeier (2018).
31. For ontological logicism, Frege requires a notion of explicit definition on which they play not only a linguistic role of introducing terms into the language for the purposes of proof, but also a metaphysical role of isolating or identifying entities (concepts or objects) of the universe. Semantically, this is expressed by specifying that the *definiendum* and the *definiens* have the same reference; cf. *Grundgesetze* §27. Importantly, it is a consistent feature of Frege's notion of definition that identification by definition is in no way

creative; they do not bring anything into existence. Existence for Frege is prior to definition. For discussion, see Boddy (2018).

32. Frege (1884, p. 102).
33. There is an exception, the value-range operator, which is a function, and so strictly speaking has no reference, as it does not have a reference in the conceptual hierarchy (it is not a concept). In *Grundgesetze* §31 Frege argues that it is nevertheless sufficient to establish that this function has a reference because value-range *terms* have reference in the conceptual hierarchy to 0-level entities. Notoriously, this argument fails. The important point here is that the logic of *Grundgesetze* is not compositional with respect to non-conceptual functions; there is no functional hierarchy playing a semantic role comparable to that of the conceptual hierarchy.
34. Frege's mistake is to specify the 0-level as entities that depend on higher levels of the hierarchy for their existence: Without concepts, there are no value-ranges. But there can only be concepts if there are 0-level entities, that is, value-ranges. The circularity is obvious.
35. Of course, in deriving the laws of arithmetic, Frege derives many theorems that are truths of value-ranges, and not specifically of numbers. But their role is only to contribute to the enterprise; their derivation is not the point of the enterprise.
36. Note that the argument is *not* that descriptive logicism is incapable of supporting the proofs of the basic laws of arithmetic, and that ontological logicism is required to satisfy this goal. Whether Frege realized this or not, we know today that there are versions of descriptive logicism that are sufficient, based on the observation that Peano Arithmetic can be derived in second-order logic with Hume's Principle as its sole non-logical axiom; the so-called "Frege's Theorem". (See Wright 1983; Boolos 1998a, c; Heck 2012 for discussion.) Actualizing descriptive logicism, however, apparently requires that the definition of number be implicit, and so rejecting Frege's insistence on explicit definition. This is particularly apparent in the context of the neo-logicist program; see Wright (1999, 2016) and May and Wehmeier (2018) on this point.
37. Frege (1979f, p. 244).
38. For Frege, whether non-Euclidian geometry is a science is up for grabs, for what would be the subject-matter if it is not about the entities given to us by intuition? It is not that it is incoherent to entertain the negation of the parallel postulate; indeed, by doing so we may learn something about the Euclidian axioms (viz. that the parallel postulate is independent of the others). But from this it does not follow that the system containing the negation of the parallel postulate is itself a science: The negation of the parallel postulate would not be an axiom if it does not govern any subject-matter. See Frege's discussion of the matter in *Grundlagen* §14.

Also relevant is Frege's discussion with Hilbert on the nature of axioms (Frege 1971); for discussion see Antonelli and May (2000).

39. As Frege puts it in "The Thought": "The thought, in itself imperceptible by the senses, gets clothed in the perceptible garb of a sentence, and thereby we are enabled to grasp it. We say that a sentence *expresses* a thought" (Frege 1977c, p. 5).
40. This remark raises the issue of whether Basic Law V constitutes an extension of logic; see fn. 2. In the Introduction to *Grundgesetze* (p. VII), Frege argues that it is not, rather it was just overlooked. Frege remarks that this neglect might be grounds for skepticism regarding its logical status, but it is a qualm that he himself had rejected.
41. Frege (1984b, p. 249). Thanks are due to Ed Zalta for bringing this passage to my attention.
42. For Frege, the notion of function is the foundational notion of logic, and it is analytic of being a function that it has a value-range: To grasp logic is thus to grasp the notion of function, and hence that there are value-ranges. They are objectually individuated by Basic Law V.
43. See Frege (1979i). In this essay, Frege speaks of the perceptual source of knowledge, rather than the empirical, as the basis of the subject-matter of the natural sciences.
44. Frege (1984a, p. 112). Frege makes the same point in *Grundlagen*, if more tersely: "The truths of arithmetic govern all that is numerable. This is the widest domain of all; for to it belong not only the actual, not only the intuitable, but everything thinkable. Should not the laws of number, then be connected very intimately with the laws of thought?" (Frege 1884, p. 21).
45. Frege (1884, p. 26).
46. In the preceeding, by "logic" I have been presuming the system of *Grundgesetze*, with its six Basic Laws. But another interpretation is possible, which is to take the logic of *Grundgesetze* itself as an applied logic, pure logic being the system as given in *Begriffsschrift*. The idea would then be that the applied logic has value-ranges as its subject-matter, and that Basic Law V is the characteristic axiom of the application that is added to the laws of pure logic. Call this the logical application. On this view, arithmetic would be a sub-application, what results from adding the definition of number (and other relevant definitions), but no further axioms, to the logical application. This would satisfy the tenets of logicism because the subject-matter of the application is comprised of logical objects, the value-ranges of functions. In effect, the logical application would be a naive set theory. Non-naive set theory could also be looked at as an application of logic, with the addition of ZF axioms and an iterative conception of set; arithmetic would be a sub-application of this theory. But we would be hard-pressed to label this the logical application. Since

non-naive sets are not analytically connected to any foundational logical notions, they would not be, on Frege's view, logical objects. So whatever we may think of the reduction of arithmetic to non-naive set theory, it would not be, as Frege sees things, a reduction to logic.

47. Frege (1893/1903, p. 9). In other places, for instance in "Function and Concept", Frege uses the term "assertion stroke". These usages are interchangeable.
48. Frege (1977b, p. 44; 1979g, p. 253; 1979e, p. 139; 1979i, p. 267), respectively.
49. Bear in mind that the details of the judging—how it was made, when or by whom—are not encapsulated in the judgement, and so not transmitted by assertions.
50. Frege distinguishes an assertion—the overt manifestation of a judgement for communication—from its being communicated with assertoric force: "Judgement is made manifest by a sentence uttered with assertive force" (Frege 1977a, p. 57). Assertoric force may be assigned to thoughts regardless of whether they are judgements; accordingly an actor can utter lines with assertoric force, but this is not to make an assertion, as the utterances are of fiction, and are neither true nor false.
51. Frege (1977b, pp. 37, 43; 1980, p. 78).
52. Frege (1893/1903, p. X).
53. A thought not paired with a truth-value is thus unjudgeable; the thought alone is not a judgeable content. Such thoughts are not proper thoughts; they are "mock" thoughts which in certain contexts, *e.g.* fiction, can be thought of as if they were judgeable.
54. Frege (1979c, p. 185).
55. Frege (1979g, p. 253).
56. Frege (1970b, p. 65).
57. Frege (1979d, p. 7).
58. Frege (1979e). I have adopted here the translation of this remark in Textor (2010).
59. Frege (1979a, p. 189). Almost the same remark is found in Frege (1979c, p. 185). These unpublished (and unfinished) writings are contemporaneous, dated to 1906.
60. Frege (1979g, p. 253).
61. Cf. Frege (1979g, p. 253).
62. There is a question as to where belief fits into this picture. Belief is like thinking in that it is not limited to true thoughts, but unlike merely thinking a thought, believing a thought involves a commitment (perhaps to some degree) to its judgeability: We believe thoughts to be true. But is belief, like thinking, anticipatory for judgement? Must we hold a thought to be true in order to judge that it is true? One would think not. Surely,

one can inquire of a thought whether it is true without hedging whether it or its negation is true. Agnosticism is no bar to judgement. But if so, all that is required to make a judgement is to think a thought, to grasp it. One may also believe it, but this attitude is ancillary to judgement.

63. See Textor (2010) for an illuminating discussion. Textor observes that taking judging to be a matter of choice does not in itself imply that this is an objectual choice—figuring out that a thought is true does not entail a commitment to the True and the False. But taking acknowledgment to be ontic, Textor argues, forces this commitment.
64. Errors of judging are cognitive errors. Agents may think they have made a judgement, but if they are mistaken, then they have failed to make a judgement, and even if they believe a thought, they could not be said to know that thought. Logical errors are of a different kind, on the order of over-sights of calculation; a proper proof cannot contain any errors, by definition.
65. See the Preface to *Grundgesetze* (p. XX*ff*) for one prominent instance of the critique of psychologistic logic.
66. Frege (1979d, p. 3).
67. Frege (1884, p. 101).
68. Frege (1884, p. 101). Frege makes this point specifically mentioning definitions, but it extends directly to axioms.
69. In *Begriffsschrift*, this is the sole inference rule; it is among others in *Grundgesetze*, but this does not undermine its centrality. The addition of the other rules, Frege notes, is conservative over the class of proofs, and serves only for the perspicuity of proofs: "This can be done," he notes, "without loosening any link in the chain of deduction, and it is possible to achieve in this way a remarkable degree of compression" (Frege 1884, p. 104, fn. 1).
70. Frege (1884, p. 2). These interconnections can be precisely mapped, given Frege's insistence that proof be gap free. Frege does this for *Begriffsschrift,* providing a chart of dependencies for the proofs therein, but not for the vastly more complicated structure of *Grundgesetze*. Heck (2012), Sect. 12.4 partially rectifies this, providing a depiction of the proof relations of some of the major theorems.
71. At this juncture, Wittgenstein in the *Tractatus* departs from Frege. While Wittgenstein agrees that logic reveals inferential dependencies, he rejects that logic is axiomatic. As noted above, no proposition is logically prior to any other on Wittgenstein's view.
72. Frege (1979d, p. 3).
73. See Frege's "Table of The Basic Laws" and "Table of Definitions" in the Appendices of *Grundgesetze* I, pp. 239–241. Also see Cook (2013) for explication of Frege's principles.

74. See *Begriffsschrift* §24 and *Grundgesetze* §27.
75. If we may add arbitrary judgements to the core in this way, we nevertheless should be judicious. It would be otiose to introduce definitions that had no worth in the derivation of a body of judgements. The fruitfulness of a definition is holistic; it resides in the role it plays in establishing a pattern of interconnected judgements by proof. For instance, the worth of the definition of number results in part from the fact that the entities so defined can be proven to form a natural sequence. See Boddy (2018) for recent discussion of Frege's view of definitions.
76. Frege (1893/1903, p. 34).
77. Key to Frege's reasoning here is the obvious principle that each occurrence of a particular symbol stands for the same thing; however one is instantiated, so must the others. This precludes one occurrence of Γ being the True, and the other the False. If this were allowed, then there would not be logical grounds for advancing to the judgement. (Let Δ and the higher occurrence of Γ be the True, while the lower occurrence of Γ is the False.) The point here, which is formal and syntactic, and not semantic, is known (without reflection) by any competent user of the Begriffsschrift.
78. The rub of course is Basic Law V. In this case we have the judgement because "a value-range equality can always be converted into the generality of an equality, and vice versa" (Frege 1893/1903, p. 36). The problem here is not with the transition to a judgement from this stipulation, rather, it lies with the stipulation itself. For each of the other Basic Laws, the stipulations that ground their judgement secure references for the relevant function terms. But not for Basic Law V; the stipulation on which it is based does not secure the reference of the value-range operator, Frege's efforts, especially in §10 and §§29–31, notwithstanding. Cf. Heck (2012), part I for a detailed discussion of just why not.
79. Conant (1991, p. 140).
80. Frege (1884, p. 4). Note that Frege's remark here is not meant to be absolute; it is not as if there is a property of propositions that must be satisfied in order to qualify as a Basic Law (for instance, as Burge (2005) holds; cf. 325*ff*). The point is that general laws do not admit of proof in the context of a particular nexus of proofs; in that context they are foundational, and depend on no other judgements. But in other contexts they may admit of proof, and which system to choose is ultimately a pragmatic choice, although a choice must be made: In a given system, no judgement can both found and be dependent. Frege says this (Frege 1893/1903, pp. 205–206):

> Whether a truth is an axiom depends therefore on the system, and it is possible for a truth to be an axiom in one system and

> not in another. … We can see from this that the possibility of one system does not necessarily rule out the possibility of an alternative system, and that we may have a choice between different systems. So it is really only relative to a particular system that one can speak of something as an axiom.

Here Frege uses the term axiom inclusively of the Basic Laws. Accordingly, it is not to be excluded that there may be other judgements that meet the criterion of being logically judged and that may be basic. If so, we will get a different system of dependencies among judgements, in which the Basic Laws may be theorems, although this variance would be conservative, since as a whole we would still have the same body of judgements. A choice is to be made; one system may be more revealing about certain concepts. may lead to more perspicous proofs, etc. But what Frege does not do is place more stringent requirements on qualifying as a Basic Law, for instance, that they can be judgements that can be only foundational, and absolutely cannot depend on other judgements.

81. Frege (1979h, p. 168).
82. Note that while Frege is employing the semantics of truth-functions as grounds for making judgements, it would not be quite accurate to say that he gives a semantic *proof* of the Basic Laws, even if the argument is inferential in a broad sense. The notion of proof is reserved for rigorous logical derivation from judgements to judgements in the Begriffsschrift. It would be to run roughshod over the very distinction that Frege is trying to make between judging and proving; it would be to confuse terminology. Also, one should be careful not to mistake Frege's argument for an elucidation, yet a third epistemic procedure that Frege calls out. Along these lines, Weiner (2014) argues that in §18 Frege is giving an elucidation of what constitutes being a logical law. But Frege's descriptions in that section says nothing about the role that judgement plays within the logical system, which is his characteristic way of specifying elucidations. Rather, what Frege says is something about the relation between a judgeable content and a judgement, the characteristic relation of judging.
83. Jeshion (2001) seems to ascribe this view to Frege. The view is problematic: If being being self-evidentially true is a *property* of thoughts, the question arises of what it is to have this property, to fall under this concept? What would it be to make the judgement of the thought that the thought that p is self-evident? Whatever the answer, it had better not be that it is self-evident, at pain of entering a regress. Also, see fn. 4.

84. Frege (1893/1903, p. 591). This remark is part of the following often cited remark: "The matter concerns my basic law (V). I have never concealed from myself that it is not as obvious as the others nor as obvious as must properly be required of a logical law." See discussion in fn. 4. In this passage, "*einleuchten*" is translated as "obvious", not as "self-evident". For reasons for preferring this translation, see Rossberg and Ebert (2013, p. xxii).
85. The notion of obviousness or self-evidence of logical laws at play here is to be distinguished from another notion. On this latter notion, all we need to become aware of the truth of a thought is to grasp that thought; no further cogitation is required, so that grasp gives us direct access to knowledge. (Reasonably, propositions of the form $a=a$ are like this.) To the extent that we label this self-evidence, it is a property of the content of a thought, not a property of judging that thought as such. The distinction here is akin to that drawn by Jeshion (2001), although she does not draw the connection to judging Basic Laws.
86. Standardly, these objects will the the True and the False, which, as Frege argues in *Grundgesetze* §10 can be taken to be logical objects; to wit, value-ranges. The conditional concept, however, is not limited to these two objects as arguments. In §12, Frege introduces the concept by specifying that its value will be the False if the antecedent is the True and the consequent is "any object that is not the True", and will the True otherwise. So strictly speaking, Frege does not define a truth-function unless the domain of objects consists of only the True and the False. This is also true of the other logical concepts.
87. Keeping in mind that judging is a psychological act, and so the epistemic notions are inapplicable. If the they apply at all, they apply to judgements.
88. Moreover, Frege would not consent to the claim that all scientific judgements are synthetic, or for that matter that some are synthetic, and others analytic. His point is quite that the notion of judgement understood in scientific contexts is orthogonal to whether they are analytic or synthetic.

References

Antonelli, A., and R. May. 2000. Frege's New Science. *Notre Dame Journal of Formal Logic* 41: 242–270.

Boddy, R. 2018. Frege's Unification. To appear in *History and Philosophy of Logic*.

Boolos, G. 1998a. On the Proof of Frege's Theorem. In *Logic, Logic and Logic*. Cambridge, MA: Harvard University Press.

Boolos, G. 1998b. Reading the *Begriffsschrift*. In *Logic, Logic and Logic*. Cambridge, MA: Harvard University Press.

Boolos, G. 1998c. The Consistency of Frege's Foundations of Arithmetic. In *Logic, Logic and Logic*. Cambridge, MA: Harvard University Press.

Burge, T. 2005. Frege on Knowing the Foundation. In *Truth, Thought, Reason: Essays on Frege*. Oxford: Oxford University Press.

Conant, J. 1991. The Search for Logically Alien Thought: Descartes, Kant, Frege, and the *Tractatus*. *Philosophical Topics* 20: 115–180.

Cook, R.T. 2013. How to Read *Grundgesetze*. In *Gottlob Frege: Basic Laws of Arithmetic*, ed. P.A. Ebert and M. Rossberg. Oxford: Oxford University Press.

Floyd, J. 2001. Number and Ascriptions of Number in Wittgenstein's *Tractatus*. In *Future Pasts: The Analytic Tradition in Twentieth Century Philosophy*, ed. J. Floyd and S. Shieh. Oxford: Oxford University Press.

Frege, G. 1950 [1884]. *Die Grundlagen der Arithmetik* [The Foundations of Arithmetic], trans. J.L. Austin. Breslau: Wilhelm Koebner; Oxford: Basil Blackwell.

Frege, G. 2013 [1893/1903]. *Grundgesetze der Arithmetik* [The Basic Laws of Arithmetic], ed. P. Ebert and M. Rossberg, trans. Philip Ebert and Marcus Rossberg. Jena: Hermann Pohle; Oxford: Oxford University Press.

Frege, G. 1970a. Function and Concept. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black. Oxford: Basil Blackwell.

Frege, G. 1970b. On Sense and Reference. In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black. Oxford: Basil Blackwell.

Frege, G. 1971. *On the Foundations of Geometry and Formal Theories of Arithmetic*, ed. E.-H.W. Kluge. New Haven: Yale University Press.

Frege, G. 1972. *Conceptual Notation and Related Articles*, trans. Terrell Ward Bynum. Oxford: Oxford University Press.

Frege, G. 1977a. Compound Thoughts. In *Logical Investigations*, ed. P. Geach. Oxford: Basil Blackwell.

Frege, G. 1977b. Negation. In *Logical Investigations*, ed. P. Geach. Oxford: Basil Blackwell.

Frege, G. 1977c. The Thought. In *Logical Investigations*, ed. P. Geach. Oxford: Basil Blackwell.

Frege, G. 1979a. A Brief Survey of My Logical Doctrines. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.

Frege, G. 1979b. Boole's Logical Calculus and the Concept-Script. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.

Frege, G. 1979c. Introduction to Logic. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.

Frege, G. 1979d. Logic (1879/1891). In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.

Frege, G. 1979e. Logic (1897). In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.

Frege, G. 1979f. Logic in Mathematics. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.

Frege, G. 1979g. Notes for Ludwig Darmstaedter. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.
Frege, G. 1979h. On Euclidian Geometry. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.
Frege, G. 1979i. Sources of Knowledge of Mathematics and the Mathematical Natural Sciences. In *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach. Oxford: Basil Blackwell.
Frege, G. 1980. Letter to Jourdain (undated; *c.* 1/1914). In *Philosophical and Mathematical Correspondence*, ed. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart. Oxford: Basil Blackwell.
Frege, G. 1984a. On Formal Theories of Arithmetic. In *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuiness. Oxford: Basil Blackwell.
Frege, G. 1984b. On Mr. H. Schubert's Numbers. In *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuiness. Oxford: Basil Blackwell.
Goldfarb, W. 2001. Frege's Conception of Logic. In *Future Pasts: The Analytic Tradition in Twentieth Century Philosophy*, ed. J. Floyd and S. Shieh. Oxford: Oxford University Press.
Heck, R.G. 2012. *Reading Frege's Grundgezetze*. Oxford: Oxford University Press.
Heck, R., and R. May. 2013. The Function Is Unsaturated. In *The Oxford Handbook of the History of Analytic Philosophy*, ed. M. Beaney. Oxford: Oxford University Press.
Heck, R., and R. May. 2018. Truth in Frege. In *The Oxford Handbook of Truth*, ed. M. Glanzberg. Oxford: Oxford University Press.
Jeshion, R. 2001. Frege's Notions of Self-Evidence. *Mind* 110: 937–976.
May, R., and K. Wehmeier. 2018. The Proof of Hume's Principle. In *Essays on Frege's Basic Laws of Arithmetic*, ed. P. Ebert and M. Rossberg. Oxford: Oxford University Press.
Rossberg, M., and P. Ebert. 2013. Translators' Introduction. In *Basic Laws of Arithmetic (Grundgesetze der Arithmetik)*. Oxford: Oxford University Press.
Textor, M. 2010. Frege on Judging as Acknowledging the Truth. *Mind* 119: 615–655.
Weiner, J. 1980. *Frege in Perspective*. Ithaca and London: Cornell University Press.
Weiner, J. 2014. The Epistemological Project of the Grundgesetze. Presented at Final Grundgesetze Workshop, New York University.
Wittgenstein, L. 1922. *Tractatus Logico-philosophicus*. London: Kegan Paul.
Wright, C. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
Wright, C. 1999. Is Hume's Principle Analytic? *Notre Dame Journal of Formal Logic* 40: 6–30.
Wright, C. 2016. Abstraction and Epistemic Entitlement: On the Epistemological Status of Hume's Principle. In *Abstractionism*, ed. M. Rossberg and P. Ebert. Oxford: Oxford University Press.

Thin Reference, Metaontological Minimalism and Abstraction Principles: The Prospects for Tolerant Reductionism

Andrea Sereni

1 Fregean Equivalences and Platonism

After critically discussing the views of Kant and Mill, in §62 of *Die Grundlagen der Arithmetik* Frege gives his well-known answer to the question "How are, then, numbers to be given to us, if we cannot have any ideas or intuitions of them?". Frege's answer, which is destined to inspire the so-called linguistic turn, relies on the Context Principle (*CP*) outlined in the Introduction of that work, and here formulated as "it is only in the context of a proposition that words have any meaning". As far as number words are concerned, *CP* immediately suggests that we should "define the sense of a proposition in which a number word occurs". To define the (sortal) concept of cardinal number, the statements whose sense is to be defined must provide, at least, with identity conditions for numbers: the truth conditions for identity statements

A. Sereni (✉)
Scuola Universitaria Superiore IUSS Pavia, Pavia, Italy
e-mail: andrea.sereni@iusspavia.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_7

among number terms must then be explained in terms of statements in which such terms do not occur.

The contextual definitions which Frege introduces immediately after are central to much of the contemporary debate in the philosophy of mathematics. Readers will be familiar with the debate. The purported definition of cardinal number by what is now called Hume's Principle, and cognate definitions—e.g. the definition of directions discussed in GLA §§ 63–67, and the infamous Basic Law V of *Die Grundgesetze der Arithmetik*—are universally quantified biconditional statements, whose right-hand side (RHS) asserts that a given equivalence relation holds among unproblematic entities of a certain kind, and whose left-hand side (LHS) is an identity statement where the identity sign is flanked by novel singular terms, delivered by a term-forming operator which takes (terms for) such entities as arguments:

$$\forall\alpha\forall\beta(\Sigma(\alpha) = \Sigma(\beta) \leftrightarrow \alpha \equiv \beta)$$

Each instance of such abstraction principles—as they are known today—should allow stipulating the truth conditions of novel identity statements: equipollence among concepts introduces identity between their extensions, equinumerosity between concepts introduces identity between their numbers, parallelism between lines introduces identity between their directions. Here, according to Frege, we convert the generality of an equality into an identity. If the stipulation is successful, we are warranted in having introduced a (sortal) concept for the objects the novel terms purport to refer to. If suitable evidence is available for the truth of at least one instance of the RHS of the corresponding abstraction, we are warranted in taking those terms to effectively perform reference to objects, and thus in taking such objects to exist.

If all goes well, we obtain a proof of existence of a range of new objects based on definitions plus the kind of evidence required to prove the relevant instance of the RHS. For cardinals, such evidence can be provided a priori via initially instantiating both concept-variables in the RHS of HP with the logical concept "$x \neq x$", which allows introducing the number 0 and from there to define concepts to be used as input in HP in order to define each finite cardinal.[1] Frege discarded HP as a candidate definition of the concept of cardinal number (mainly for its falling prey of the Caesar's problem[2]), and gave an explicit definition of number in terms of extensions of second-order concepts of equinumerous

concepts. Today we know that HP, when added as an axiom to system of full impredicative second order logic, gives rise to a consistent system (Frege's Arithmetic) and allows the derivation of the second-order Dedekind-Peano axioms for arithmetic. If HP counts as an acceptable definition of cardinal number, this result shows those axioms to be derivable from logic and definitions alone, i.e. analytic in Frege's sense (GLA, §3), as argued by neo-logicists. It would also prove the existence of countably many finite cardinals as an analytic consequence of Frege's Arithmetic, hence showing that numbers are objects that can be accessed epistemically by an entirely a priori route in no need of intuition or mental representations: it would eventually vindicate something akin to logicist platonism.

But does it really? Clearly, this result is in sight only if HP and cognate abstractions can score two prerequisite results: they must in fact introduce a concept, and this must be a concept under which objects fall. Those driven by nominalistic scruples are likely to reject one of these two claims. Even beyond this, a defense of neo-Fregean platonism is hostage to a proper semantic analysis of what is achieved through contextual definitions.

This is partly a matter of Fregean exegesis. When introducing contextual definitions in GLA, Frege was working, at least explicitly, with an undifferentiated notion of *Inhalt*, before the *Sinn/Bedeutung* distinction was introduced. While in *Grundgesetze*, where such distinction is fully operative, little emphasis is put on contextual definitions, the Context Principle is no clearly in play, and it also appears in sharp tension with the compositional semantics developed by Frege.

But the question of what contextual definitions can achieve has a theoretical interest on its own, and the discussion of abstraction principles has witnessed a variety of interpretations. One can easily identify the two extremes in the spectrum of available interpretations. On one extreme stands the view already sketched: HP is an implicit definition of the sortal concept of cardinal number, under which abstract objects fall; number words qualify as genuine singular terms, and any evidence for the truth of an identity statement in which they occur is also evidence for their having genuine objectual reference and for the existence of their referents. On the other extreme, a radical reductionist attitude would take such principles as showing that statements allegedly about a range of abstract objects can in fact be paraphrased away as claims about concepts: far from supporting platonism, the equivalences deliver a reductive

translation procedure for purifying our language from the apparent reference to numbers.

Three concerns opens up here. First, on what basis can it be maintained that the proper explanatory direction of an abstraction principle goes one way or the other. Second, whether these two extreme views really exhaust the available interpretations of contextual definitions, or whether intermediate views are available. Third, supposing this is so, how such intermediate views can be motivated and developed.

In what follows, after rehearsing the opposition between the two extremes, I will investigate how intermediate views can be argued for. I will thus explore how different such views, while steering clear from radical nominalist readings, can mitigate the robust platonism of Fregean inspiration, and claim that prospects for such intermediate views are still in sight without betraying the essential insights inherited from Frege.

2 Intolerant Reductionism and Robust Abstractionism

Michael Dummett (1991, Chapter 15) identifies at least two different readings of equivalences such as HP or the Direction Equivalence. *Intolerant Reductionism* has it that

> to explain sentences about directions by translating them into sentences about lines, or sentences about numbers by translating them into ones involving expressions for and quantification over second-level concepts, commits us to denying that there are any such objects as directions or numbers. (p. 191)

Intolerant reductionists oppose all readings of the equivalences that attribute reference to the terms in the LHS.[3] Such reductionist must then explain how it is that the surface syntax of the statements introduced by the equivalences "precludes discernment of any genuine semantic structure in the sentences so explained" (ibid.). Two different paths are at her disposal. She can acknowledge that the syntactic structure of the LHS is that of an identity statements, but deny this genuinely reflects, at the semantic level, a claim about the referents of the corresponding terms. A more radical kind of intolerant reductionist can take the equivalences as explicit definitions of the expressions in their LHS's and conceive of them holophrastically, as unstructured wholes devoid of *any* syntactic pattern structure: what seem to be meaningful words occur

in the expressions defined just as "pun" occurs in "punishment". The burden of proof will be evidently more pressing for the latter kind of reductionist.[4]

Opposite to Intolerant Reductionism stands what I shall call *Robust Abstractionism*, championed by neo-logicists:

> the contextual definitions succeed in conferring upon sentences containing the terms contextually defined senses which warrant our viewing them as having just that semantic structure which their surface forms suggest. The defined terms are genuine singular terms, with a genuine reference, albeit to abstract objects. (ibid.)

Robust Abstractionism relies on an interpretation of Frege's Context Principle which neo-logicists qualify as the *Syntactic Priority Thesis*:

> According to this thesis, the question whether a particular expression is a candidate to refer to an object is entirely a matter of the sort of syntactic role which it plays in whole sentences. If it plays that sort of role, then the truth of appropriate sentences in which it so features will be sufficient to confer on it an objectual reference; and questions concerning the character of its reference should then be addressed by philosophical reflection on the truth conditions of sentences of the appropriate kind.[5]

Intolerant Reductionism and Robust Abstractionism correspond, respectively, to what Wright (1983) calls an *austere* and a *robust* reading of the equivalences such as HP. Dummett and the neo-logicists, however, diverge on the exhaustiveness of this twofold taxonomy. The question is whether there is any coherent option occupying an *intermediate* position. Dummett believes that such intermediate views exist, and argues in favor of one. Neo-logicists deny that Dummett's suggestion is a viable alternative. But also a number of other recent views, as we shall see, can be players in this game.

3 Intermediate Views: The Best of Both Words?

Any alternative to Robust Abstractionism will entail that some ingredient of the neo-logicist reading of abstraction principles must go. But which ones? Two different intuitions pull in different directions here. On the one side, one can acknowledge ascription of reference to objects to the

terms defined but offer an analysis of what it takes to be such objects which defuses the epistemic concerns elicited by standard conceptions of platonism. On the other side, one can deny that Fregean equivalences are actually capable of ascribing to the new terms the sort of referential role which is fact required by platonism. Accordingly, one can either suggest that no referential role at all is ascribed to those terms, or that only a weak form of reference can be so ascribed.

In a slogan, then, the two strategies for an intermediate view point in two different directions: either we accept the abstractionist semantics, and defuse its metaphysical import; or we reject the metaphysical import by defusing the semantics. Let us review some proposals by exploring either these paths.

Metaontological Minimalism and Thin Objects

Ontology is that part of philosophy that, following Quine, studies what there is.

Metaontology is a higher-level study of the central notions and methodology of ontology.[6] It is not concerned with establishing what objects there are, but rather with clarifying what it takes to be an object. Recently, a few authors have defended *metaontological minimalism*, which Linnebo (2012) defines as the claim that "there are objects whose existence does not (loosely speaking) impose a very demanding requirement on reality" (p. 139). More precisely, it is "the claim is that *our concept of an object* is such that it allows for thin objects", namely objects "whose existence does not impose a very demanding requirement on reality" (p. 140). Linnebo mentions physical particles as an example of thick objects. By elaborating a vivid metaphor introduced by Rayo (2013, p. 4; cf. also Rayo 2013), if God had to add particles to the inventory of existing objects, he would have had to actually do something more than what was required to be done to create the world up to that point; on the contrary, thin objects are such that, given a certain state of the world, there wouldn't be anything "extra" that God would have to make to bring such objects into existence. As Linnebo (2012, p. 140) puts it, "an object x is thin relative to some other objects if, given the existence of these other objects, very little is required for the existence of x". Borrowing a terminology from a (still) independent debate concerning metaphysical grounding (here understood as a relation of non-causal metaphysical explanation), and obliterating the difference

between "very little" and "nothing", one could say, with Shaffer (2009), that thin objects are "free lunches": they are metaphysically derivative objects whose existence is granted, as it were, for free once the existence of some appropriate range of fundamental objects is secured.[7]

Linnebo (2012, pp. 141–142) stresses the fact that the appeal to thin objects—despite having the "tendency to come across as a piece of philosophical magic that aspires to produce something out of nothing, or much out of little"—can help assuage a number of traditional worries: the problem of epistemic access to an ontology of abstract objects—as emphasized by Benacerraf (1973)—and the problem of how we are justified in postulating the existence of an infinity of such objects.

Some candidate forms of metaontological minimalism are not relevant for our present purposes. Linnebo discusses Coherentist Minimalism, i.e. "the view that the coherence of a mathematical theory suffices for the existence of the objects that the theory purports to describe" (p. 143), as Hilbert suggested in a well-known letter to Frege. Frege notoriously opposed Hilbert's view. In contemporary terms, the essential traits of the Frege-Hilbert controversy can be found in two related contexts: the neo-logicist distinction between innocent and arrogant definitions—where a definition counts as arrogant if additional epistemic work beyond linguistic stipulation is needed in order to establish the truth of the stipulated statement (cf. Hale and Wright 2000, p. 121; Hale and Wright 2009, pp. 289–290, 315); and the objection, levelled by neo-logicists against Shapiro's *ante rem* structuralism (cf. Hale and Wright 2002b), according to which the axiomatic characterization of a structure can at best provide us with a concept of a given structure but not with any existing instance of that concept. Since the intermediate views being investigated here are likely to agree with such Fregean rejoinders, coherentist minimalism will not be relevant to our discussion. More controversially, we will abstain from considering neo-logicist abstractionism as a form of metaontological minimalism. True, as Linnebo stresses, Hale and Wright have repeatedly emphasized that "the recognition that the truth of the right-hand side of an instance of a good abstraction is *conceptually* sufficient for the truth of the left" (cf. Hale and Wright 2009b, p. 193), and this apparently goes in the minimalist direction. However, in replying to Eklund (2006)'s suggestion that the neo-logicist reading of abstractions can only be defended provided it is backed by a metaphysical view like maximalism—i.e. "the thesis that whatever *can* exist *does*", or more to the point that "for any sort or kind

of objects F, if it is *possible* that Fs should exist, they *do*"—neo-logicists have hastened to make clear that on their reading of abstraction principles, the above-mentioned conceptual sufficiency entails that "there is no gap for metaphysics to plug, and in that sense no 'metaontology' to supply" (ibid.).[8] As we see it, the basic idea behind this is that the sort of linguistic and conceptual machinery which is involved in the right-to-left direction of good abstraction principles is enough to introduce genuine singular terms apt for objectual reference (and that, under favourable conditions, the effectiveness of such reference can be proved once the truth of an instance of an abstraction can be independently secured), without any need of presupposing any substantial metaontological stance. On its face, this tells against both metaontological maximalism and minimalism. Moreover, the kind of platonism defended by neo-logicists appears much closer to Frege's own, and unconcerned with qualms about thickness or thinness: abstractions provide the means for referring to abstract objects which exist in the unique sense in which everything which exists can be said to exists. Robust Abstractionism, therefore, stands out as much more congenial to a rather robust metaphysical picture of the arithmetical domain. On a plausible understanding, that picture entails that whatever requests must be made of the world in order to grant the existence of such objects, they will be as thick as those for any other (abstract or concrete) existent objects. Rather than offering an epistemology in need of thin objects, the abstractive strategy seems rather to offer, via considerations of meaning, a thin epistemology for thick objects.

Two recent developments of Fregean strategies would be more relevant here: Agustin Rayo's trivialism, and Linnebo's defence of thin objects.

According to Rayo (2013; cf. also 2014), there is a class of 'just-is' statements which is able to account for the meaning of arithmetical statements moving from an understanding of statements formulated in non-arithmetical language. The basic idea is that, e.g. for the number of dinosaurs to be 0 *just is* for there to be no dinosaurs, and, more generally:

> [NUMBERS] for the number of the *F*s to be *n* just is for it to be the case that $\exists!_n x(Fx)$.

Rayo discusses several strategies for doing away with platonist commitments, and find all of them wanting. He then offers a strategy for

specifying the nominalistic content of arithmetical statements which does not rely on paraphrases. This method allows finding a "trivialist" semantics for arithmetical statements which is not nominalistic. The trivialist stands out as a platonist. One who accepts [NUMBERS] accepts the apparently simple idea that "there is *no difference* between there being no dinosaurs and their number's being zero, in the same sort of way that there is no difference between drinking a glass of water and drinking a glass of H_2O". As an "immediate consequence", "a world without numbers would be *inconsistent*". According to trivialism, numbers exist, but their existence is "a trivial affair": once certain facts that can be nominalistically described hold, there is nothing more which is required for the existence of a suitable range of abstract objects—e.g. nothing more is required for the number of forks and knives on the table to be identical and to exist than the existence of a bijection between the forks and the knives. The creation metaphor has its grip here: nothing extra must be done by God to create numbers once he creates concepts standing in appropriate equinumerosity relations. As also Linnebo (2012, p. 148) points out, more robust forms of platonism would accept as totally intelligible the possibility of having a world in which "there should be eight planets but no object—namely the number eight—that numbers the planets". Such robust platonist won't take the existence of numbers as something we may get for free once non-arithmetical facts holds. It is thus the metaontological claim that non-thick objects are possible that underlies Rayo's trivialist platonism.

Linnebo has developed a theory of thin objects in support of a non-robust form of platonism based on abstraction principles. A full assessment must await a proper analysis of a recently published work (cf. Linnebo 2018), but the basic ideas will suffice here. Linnebo considers how, via abstraction principles, we can extend a previously available language through the introduction of newly defined terms, e.g. terms for directions. This procedure has some key ingredients. First, statements on the RHS of an abstraction (e.g. statements of parallelism about lines) will provide assertibility conditions for statements on its LHS (statements about directions). Second, statements in the LHS will behave inferentially exactly like statements which genuinely are about the purported referents of the introduced singular terms (directions). According to Linnebo, these two desiderata can be secured via two concurring thoughts. Abstraction principles are seen as means for delivering identity conditions for objects (e.g. identity conditions for directions in terms of

parallelism about lines), and reference is secured once criteria of identity are available, independently of other more direct or quasi-perceptual methods. Easy reference entails easy being, in so far as the problem of establishing what objects there are is in fact conceived as a question concerning which terms have reference. This gives rise to a form of platonism where numbers stand out as thin objects. Again, it is metaontological minimalism which paves the way for the development of a view which is Fregean in spirit, but steers clear of the robust metaphysical picture traditionally linked with platonism.

Linnebo focuses on an asymmetric aspect of abstraction principles. The right-to-left direction expresses a sufficiency condition, i.e. that the truth of statements in the RHS suffices for the truth of statement in the LHS. This entails a form of reconceptualization of the RHS into the LHS. This unidirectionality avoids some of the problems encountered by Frege's notion of "content-recarving". Frege (GLA, §62) claims that in "removing what is specific" in the content of the equivalence relation on the RHS and in "replacing it with the generic symbol =" on the LHS, "we carve up the content in a way different from the original, and this yields us a new concept". Content-recarving faces a severe problem. The biconditional formulation of abstraction principles entails a symmetric relation between its two sides: the truth conditions of the RHS are supposed to be the same as those of the LHS. In other terms, it is the same content that is expressed—the same fact that is described—on both sides of an abstraction and merely gets different descriptions. But how is this possible? How is it possible that a certain fact (e.g. parallelism between lines, equinumerosity between concepts) can be described without any mention of a given range of objects (directions, numbers) even though those objects are essential constituents of that very fact? As we shall see, this problem lurks also behind the attempts at recovering an intermediate reading of abstraction not by assuming some specific metaontological views, but by modifying the notion of reference that such principles would be able to license.

Tolerant Reductionism and Thin Reference

Dummett believes that a view intermediate between Intolerant Reductionism and Robust Abstractionism can be characterized as Tolerant Reductionism. The tolerant reductionist

> holds that the contextual definition serves to explain what it means to say, 'There is a direction orthogonal to those of line *a* and *b*', or 'There is a prime that divides both 943 and 1357', rather than to show that we ought not to say things of that kind. He therefore agrees that it would be wrong to say that neither directions nor numbers exist, even if we adopt the contextual definitions: you cannot consistently combine the assertion that there is a number satisfying a certain condition with the declaration that there are no numbers whatever. He recognises further that '"31" refers to an object' can be construed untendentiously as simply the equivalent, in the formal mode, of 'There is such a number as 31', and hence as uncontroversially true. What he denies, however, is that the notion of reference, as so used, is to be understood realistically. (p. 191)

In short: while the intolerant reductionist fails to recognize that nothing more is needed for a given class of singular terms to possess reference than they should occur in suitably selected true statements, even if we are not capable to indicate ostensively the objects referred to, those who adopt the robust interpretation err in believing that reference is bestowed on the contextually defined terms in a *robust* sense, as opposed to a *thin* sense, as Dummett calls it.

Admittedly, the distinction between thin and robust reference needs be spelled out properly. As Picardi (2016, pp. 34–35) puts it:

> as it stands, the contrast between a thin and a robust conception of reference in the case of number words also appears elusive. Indeed, if, following the suggestion put forward by Dummett in *The Interpretation of Frege's Philosophy*, we construe Frege's notion of *Bedeutung* as semantic value, and ascribe a semantic value to incomplete expressions, there seems to be little point in denying a semantic value also to names of abstract objects. But appearances are deceptive: lest we give the impression of stipulating abstract objects into existence, or of treating ascription of reference to number words as a mere *façon de parler*, more has to be said about the way in which the referents in question play an operative role in determining the truth-conditions of the relevant sentences.

Dummett's distinction between different ways in which reference might be said to be possessed by different expressions, then, goes back to his account of incomplete expressions.[9] Dummett distinguishes between *semantic role* and *realistic reference*. Semantic role is the contribution that an expression effects to the determination of the truth-conditions

of the sentences in which it occurs. Reference is, on the contrary, understood realistically when it is the relation that an expression bears to the objects it stands for. Realistic reference is shaped in analogy with the name/bearer relation that is essential to the understanding of proper names. Dummett (1973, pp. 223–224) believes that it is only reference as semantic role that can and should be attributed to incomplete expressions. And it is on this model that he believes that reference for abstract objects introduced by contextual definitions should be conceived of, as being bestowed in a thin sense, and exhausted by semantic role.[10] Grasping the thought expressed by a sentence, and thus knowing what the truth-conditions of that sentence are, requires us to determine the contribution of the component expressions. In general, we will have a "particular way of conceiving of a certain object as being picked out by each singular term", such that knowledge of the reference of the relevant singular term is essential for determining the truth-conditions of the sentence in which it occurs. But the situation is different "when the sense of a given term is given by means of a contextual definition":

> Our grasp of the thought expressed by a sentence containing the term is mediated by our knowledge (possibly only implicit) of how to arrive at an equivalent sentence not containing that term. The notion of the reference of the term, as determined by its sense, plays no role in our conception of what determines the thought as true or false, nor, therefore, in our grasp of the thought; the attribution of reference to the term may be defensible, when tolerantly viewed, but is semantically idle. (p. 193)

Now, whatever sense may be given to the idea of reference being "idle" in the equivalences for directions and numbers, the thought is evidently contrasting the neo-logicists' (once held) thesis of *Hidden Reference*,[11] i.e. the thesis that in order to function as proper explanations of the concepts introduced, the RHS's of the equivalences must be seen as *not* being innocent of commitment to the (abstract) objects falling under the concept thereby defined, even though there is no expression whose function is to overtly refer to those objects.

Dummett's tolerant reductionism and Hale's and Wright's robust platonism diverge under various respects. First, they attribute a different role to reference to abstract objects in determining the content of the statements in the RHS of the equivalence. Dummett attributes

none, Hale and Wright attribute an essential one. They thus have different conceptions of what is involved (what role reference should have) in the process of "carving up" a content. Dummett believes that reference is idle in this process; Hale and Wright have it that reference is playing an operative role, even though this role is not apparent from the surface syntax of the RHS of the equivalences. Finally, they have different conceptions of the sense in which terms for abstract objects can be said to have reference. Dummett thinks that they have only thin reference, or semantic role; Hale and Wright contend that they have robust reference.

The proper formulation of the first point of divergence is, however, a matter of some dispute. Neo-logicists acknowledge that identity of sense between the statements in the two sides of the equivalences would require reference to the relevant objects to be detectable in the surface structure of both sentences. What they question is that it is identity of *sense* that they need. Rather, they claim, the best way of making sense of the equivalences is to distinguish between identity of sense and identity of content. The latter is identified, roughly, with the sentences' truth-conditions, as neatly distinguished from the *thoughts* that the sentences express. It is only identity of content that neo-logicists need for their claim (cf. Hale 1994, p. 196). In their terminology, while the RHS of the equivalences has *epistemological* or *explanatory priority*, it is the LHS that has *ontological priority*.

There is another point of dispute. Neo-logicists deny that the distinction between thin reference (semantic role) and robust reference (realistic reference, in analogy to the name/bearer model) is tenable. They claim that once the proper import of Frege's Context Principle is understood, nothing more is required, in order for a singular term occurring in the equivalences to have (robust) reference, than it possesses semantic role. Frege's Context Principle should be seen as mandating that it cannot be a requirement for a singular term to have reference that its object can be picked out ostensively, or by causal connection, or by mental association, or by any means which is usually associated to the name/bearer relation. A singular term's having semantic role is all that is needed for construing reference to the objects of abstraction as a relation to something external. There is only one notion of reference. As it seems, exactly as there is no gap for a thin conception of objects to fit in, there is also no room for a thin conception of reference.

4 The Prospects for Intermediate Views

Dummett's misgivings against Robust Abstractionism partly rely on whether Frege's semantic views support Hale's and Wright's interpretation. As mentioned, the pivotal role that the Context Principle has in GLA is in apparent contrast with the compositionalist semantics of GGA, since compositionality seems to attribute constitutive priority to subsentential expressions in the determination of meaning. According to Dummett (1991, p. 196), the question of whether the robust reading of abstractions can be endorsed while endorsing other aspects of Frege's views in GLA

> depends on what 'robust' means. If it means simply 'not austere', so that a robust view is merely the negation of the view that a reference may not be ascribable to the new terms at all, then Wright's interpretation is indeed faithful to *Grundlagen*. But, when we recall that his formulation of the context principle involved the ascription of an 'objectual reference', we may doubt this mild reading of the word 'robust'. From the standpoint of *Grundgesetze* there is a more substantial notion of reference than the thin one allowed by *Grundlagen*, that notion, namely, employed in the semantic theory: to ask whether an expression has a reference in this sense is to ask whether the semantic theory assigns one to it, or needs to do so, where reference is a theoretical notion of that theory. Wright's notion of objectual reference appears to be just such a substantial notion, at home in *Grundgesetze* but not in Grundlagen.

The issue seems thus to reduce to establishing whether the doctrines of *Grundlagen* allow for the thin view of reference which Dummett suggests for his own intermediate view. And here there may be room for maneuver. A number of remarks in Picardi (2016, pp. 51–52) help seeing this:

> Frege's use of the word "*Inhalt*" is ambiguous both before and after 1890, and generally, when applied to singular terms, indicates the thing signified by the sign, and not the sense of the sign. In § 65 of *GLA* Frege says that Leibniz's Law offers a definition of identity […] In my opinion the appeal to Leibniz's law shows that, as far as singular terms are concerned, the distinction between *Sinn* and *Bedeutung* is already operative in *GLA*, possibly in an inchoate form, together with the new interpretation of identity, very similar to the one which he was to employ after 1890 […] In *GLA* Frege seems to take it for granted from the very beginning that proper names and number words flanking an identity sign stand for objects. […]

> What speaks in favor of my interpretation is that Frege himself, both in [*Funktion und Begriff*] and in the *Preface* to *GGA*, stresses the continuity between *GLA* and his later work. As far as the notion of *Inhalt* is concerned, also in Frege's post-1890 writings there is a lingering ambiguity […] and my surmise is that also in *GLA,* as far as singular terms are concerned, the word "*Inhalt*" is used to indicate "the thing signified", the referent of a sign. […] in *GLA* § 64 […] "*Inhalt*" refers to the content of the *judgment* "Line *a* is parallel to line *b*", which in the post-1890 terminology will be called a "thought". However, what the argument is meant to show is that the content of the identity statement "The direction of line *a* is the same as the direction of line *b*", is not, or not only, that of an analytically equivalent reformulation of the content of the first judgment (unless the judgment about lines already contains an implicit reference to directions, as Neo-Fregeans surmise). In the identity statement the semantic role of the newly introduced expressions is that of singular terms purporting to stand for the thing signified. Perhaps, as Dummett (1982) suggests, the word "direction" is here capable only of a thin construal. However, a tolerant reductionist acknowledges that a new ontological commitment is undertaken which was absent from the original judgment.

If this is on the right track, we can follow Picardi in concluding not that, as Dummett (1991) suggests against the neo-logicists, Frege's semantic ideas in GLA forbid the ascription of any robust notion of reference to singular terms, but rather that, since "in *GLA* the CP was meant to apply not only at the level of sense, but also at the level of the *thing signified*", then "in the absence of a full-fledged semantic theory, it is difficult to assess how weak or strong the notion of reference, as applied to singular terms, is required to be". GLA is thus "semantically indetermined", and open to a variety of interpretations (Picardi 2016, pp. 54–55).

The crucial question, then, is not whether it is possible to endorse Frege's views in GLA and at the same time adopt either an austere or robust reading of the definitional equivalences. That the former is ruled out in that context, and that the latter is permissible, should by now be taken as established. Rather, the interesting question is whether either metaontological or semantic consideration can themselves lend plausibility to an intermediate reading of abstractive definitions which would, as a bonus, deliver a picture of reference to abstract objects somewhat more palatable that the robust metaphysical view often associated with Fregean platonism.

We have reviewed two families of strategies for obtaining such intermediate position. Metaontological minimalists develop abstractionist insights in the framework of a general conception of what it takes to be an object, a conception which allows for the possibility of objects which impose little if no demand on reality for their existence. On the other hand, Dummett's suggested intermediate view does not depend on metaontological presuppositions, but rather investigate whether contextual definitions allow for a notion of reference short of a thick model of reference.

On this latter score, the taxonomy we have reviewed may not even be exhaustive. After all, it is as yet not completely clear whether Dummett's tolerant reductionism should count as a form of platonism or as a form of nominalism. For while Dummett's aim is to rehabilitate a face-value reading of arithmetical statements as involving reference to objects, his insistence on the thin nature of such reference may be well available to some nominalist proposals.

Hale and Wright themselves point out that Dummett's overall skepticism towards abstraction principles—which also include worries on the Bad Company objection, the impredicativity of HP, and the Caesar Problem—suggest the following *indeterministic thesis*:

> Fregean abstraction need never be taken to introduce a concept of a *new* kind of thing – and hence cannot be regarded as introducing a concept of a distinctive kind of thing at all[12]

If that is correct, Hale and Wright (ibid.) continue,

> then there is a kind of nominalist – let's call him a *semantic nominalist* – who can acknowledge the success of the Fregean explanations in the kind of cases illustrated: he can wholeheartedly accept the equivalences, without further conditions, allowing that the semantic structure of the left-hand sides is just what it seems; this while he legitimately – according to Dummett – construes the ontology of the left-hand sides as involving no excursion outside nominalistic boundaries. […] So far, then, from supplying the resources for legitimating a modest form of platonism, Fregean abstraction promises to emerge – If Dummett's indeterministic thesis is right – as a powerful tool for the rehabilitation of what appears to be talk of *abstracta* in terms of a wholly concrete ontology.

Take, for a relevant example, an understanding of abstraction principles suitable to the sort of fictionalism advanced by Yablo (2001, 2005).

Here contextual definitions could be seen as fiction-introducing principles, stating on which conditions we are entitled to take a certain range of literally false, platonistically construed statements as fictionally true. Such a fictionalist refrains from any platonist reading of number-talk as introduced by HP, but would still distance herself from nominalistic readings which see the LHS of such principle as mere rephrasing of the nominalistically acceptable RHS, or from holophrastic construals which take statements on the LHS as wholly devoid of syntactic structure.

It is far from obvious that a suitably elaborated fictionalist semantics could not exploit Dummett's insights to support a fictionalist reading of abstraction principles which is a close relative of Tolerant Reductionism.[13] This may in the end be wholly compatible with what Dummett (1991, p. 192) says about the notion of reference in GLA: "The context principle, as enunciated in *Grundlagen*, can be interpreted as saying that questions about the meaning (*Bedeutung*) of a term or class of terms are, when legitimate, internal to the language. […] any legitimate question about the meaning of a term, that is, about what we should call its reference, must be reducible to a question about the truth or otherwise of some sentence of the language". As Picardi (2016, p. 54) reminds us, "GLA seems hospitable both to Carnap's and to Wright's interpretations":

> Carnap's masterpiece, *Der logische Aufbau* der *Welt*, draws generously on Frege's and Russell's conceptions of definitions in use and by logical abstraction. Carnap construes Frege's contextual dictum as applied to class names as a means of introducing in the language names of quasi-objects. Such names do not stand for classes; rather, in the context of a sentence, indicate the quasi objects made available at a given stage by the constructional method […]. What I would like to suggest is that Carnap's construal of definitions in use (contextual definitions) of number words and class names is a legitimate interpretation of what Frege is up to in *GLA*, and can be taken as an example of the stance of the tolerant reductionist described by Dummett in chapter 15 of *FPM*, and contrasted with the austere reductionist and the robust interpretation of the notion of reference provided by Wright (1983). (Picardi 2016, p. 39)

It is not unthinkable, then, to view abstraction principles as introducing some suitable form of fictional reference.[14] And it may not be just a coincidence that fictionalist proposals on the lines just reviewed, on their turn, draw so heavily on Carnapian insights, rephrasing the

internal/external distinction as a distinction between literal and non-literal content. And yet other brands of fictionalism, based on a Russellian no-class theory, may be built upon non-platonist readings of abstraction principles.[15]

If Dummett and Picardi are right, then, the lack in GLA of a semantic theory akin to the one at play in GGA is what makes Frege's appeal to contextual definitions so much open for different interpretations. The sharp contrast between an austere and a robust reading, between Robust Abstractionism and Intolerant Reductionism, between a overly realistic conception of platonism and an unduly restrictive nominalist credo, which have dominated large part of this area of inquiry in the past, hardly exhausts the routes for achieving a tenable epistemology for mathematics grounded on the crucial role of contextual definitions and abstraction principles in particular. Be it the metaontological minimalist way based on a defense of thin objects, or the exploration of different varieties of thin reference, or yet other proposals, the prospects for an intermediate view are far from being foreclosed, and very unlikely to abandon the many paths that can be drawn from Frege's insights.[16]

Notes

1. As suggested by Frege and detailed in recent reconstructions of Frege's Theorem; cf Boolos (1987), Heck (1993, 2011), Zalta (2017).
2. Namely, the inability of eliminating numerical identities in favour of statements of equinumerosity of concepts when fed with identities of the form "The number of F = q", where "q" is not of the form, nor abbreviates terms of the form, "the number of G".
3. Intolerant Reductionism differs from *Rejectionism* as advanced by Field (1984): the rejectionist rejects the equivalence as false on grounds that there are no abstract entities, but agrees that arithmetical statements should be interpreted at face-value. See also Wright (1990).
4. For a discussion, cf. Rosen (1993).
5. Wright (1983, pp. 51–52).
6. Cf. Berto and Plebani (2015).
7. On grounding, see Correia and Scnheider (2012). Few works have so far applied the notion of grounding to the philosophy of mathematics (cf. Rosen 2010, 2011; Schwarztkopff 2011; Donaldson 2016), but a proper exploration of connections between the grounding debate and issues in the philosophy of mathematics is still lacking. Cf. also Linnebo (2018), §1.8.

8. Cf. also Ibid., fn. 8, p. 181. The doctrine Hale and Wright call "minimalism" and appeal to should (at least in their eyes) be identified with "the abstractionist metaphysics of abstract objects, and of reference to them" which they discuss and defend (ibid., p. 207), and thus as unrelated to metaontological concerns. For criticisms, cf. Eklund (2016, esp. p. 90).
9. Dummett (1973, Chapter 7).
10. Although Dummett (1991, p. 164) slightly reconsider the significance of this analogy, it seems undeniable that Dummett's tolerant reductionism, as formulated in his (1991), is reminiscent of the idea of thin reference as was attributed to incomplete expressions in his (1973).
11. The thesis is called Hidden Reference in Hale (1994). The thesis is first presented in Wright (1983, p. 31ff). A discussion is in Wright (1988). Hale and Wright seem not to have further developed this idea, claiming merely for identity of existential commitments between the two sides of the equivalences. Cf. Hale and Wright (2009b, p. 187):

    > […] the existential commitments of the statements which the abstraction pairs together *are* indeed the same—and hence the right-hand side statements, no less than the Σ-identities, implicate the existence of Σ-abstracts while containing no overt reference to them.

12. Hale and Wright (2002a, pp. 355–356).
13. Some of the present materials is taken from the author's Ph.D. dissertation (*Speaking of Abstract Objects*, University of Bologna, 2006), where, among other things, such a fictionalist reading of abstraction principles is further explored. In retrospection, I wish to thank once more Eva Picardi for her precious supervision of that work, as well as Crispin Wright for his many comments as an external reader.
14. Or, as Picardi (2012) suggests to call them in commenting Thiel's work, "'auxiliary objects', depending for their identification on the activities performed by mathematicians".
15. Cf. Florio and Leach-Krouse (2017).
16. I wish to thank the editors of the present volume for giving me the opportunity of homaging Eva Picardi's memory, her long-lasting teachings and her illuminating contributions to the study of Frege's philosophy.

References

Benacerraf, P. 1973. Mathematical Truth. *The Journal of Philosophy* 70: 661–679.
Berto, F., and M. Plebani. 2015. *Ontology and Metaontology: A Contemporary Guide*. London and New York: Bloomsbury.

Boolos, G. 1987. The Consistency of Frege's *Foundations of Arithmetic*. In *On Being and Saying: Essays in Honor of Richard Cartwright*, ed. J. Thomson, 3–20. Cambridge, MA: MIT Press.

Correia, F., and B. Schnieder (eds.). 2012. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press.

Donaldson, T. 2016. The (Metaphysical) Foundation of Arithmetic? In *Nous* 51 (4): 775–801.

Dummett, M. 1973, 1981. *Frege: Philosophy of Language*. London: Duckworth.

Dummett, M. 1991. *Frege: Philosophy of Mathematics*. Cambridge, MA: Harvard University Press.

Eklund, M. 2006. Neo-Fregean Ontology. *Philosophical Perspectives* 20: 95–121.

Eklund, M. 2016. Hale and Wright on the Metaonotology of Neo-Fregeanism. In *Abstractionism*, ed. P. Ebert and M. Rossberg, 79–93. Oxford and New York: Oxford University Press.

Field, H. 1984. Platonism for Cheap? Crispin Wright on Frege's Context Principle, originally printed as "Critical Notice of Crispin Wright: *Frege's Conception of Numbers as Objects*" In *Canadian Journal of Philosophy*, 14; reprinted in H. Field (1989), *Realism, Mathematics and Modality*, Essay 5. Oxford: Blackwell.

Florio, S., and G. Leach-Krouse. 2017. What Russell Should Have Said to Burali-Forti, *Review of Symbolic Logic* 10 (4): 682–718.

Frege, G. (GLA), 1884. *Die Grundlagen der Arithmetik: eine logische mathematische Untersuchung über den Begriff der Zahl*. W. Koebner, English translation by J.L. Austin (1953). *The Foundations of Arithmetic*. Oxford: Blackwell.

Frege, G. (GGA), 1893–1903. *Die Grundgesetze der Arithmetik*, 2 vols. Jena: H. Pohle.

Hale, B. 1994. Dummett's Critique of Wright's Attempt to Resuscitate Frege. *Philosophia Mathematica* 3 (2): 122–147.

Hale, B., and C. Wright. 2000. Implicit Definition and the a Priori. In *New Essays on the a Priori*, ed. P. Boghossian and C. Peacocke, 286–319. Oxford and New York: Oxford University Press.

Hale, B., and C. Wright. 2002a. To Bury Caesar…. In *The Reason's Proper Study*, ed. B. Hale and C. Wright. Oxford: Oxford University Press.

Hale, B., and C. Wright. 2002b. Benacerraf's Dilemma Revisited. *European Journal of Philosophy* 10: 101–129.

Hale, B., and C. Wright. 2009a. Focus Restored: Comments on John MacFarlane. *Synthese* 170 (3): 457–482.

Hale, B., and C. Wright. 2009b. The Metaontology of Abstraction. In *Metametaphysics*, ed. D. Chalmers, D. Manley, and R. Wasserman. 178–212. Oxford: Oxford University Press.

Heck Jr., R. 1993. The Development of Arithmetic in Frege's Grundgesetze der Arithmetic. *Journal of Symbolic Logic* 58: 579–601.

Heck Jr., R. 2011. *Frege's Theorem*. Oxford and New York: Oxford University Press.

Linnebo, Ø. 2012. Metaontological Minimalism. *Philosophy Compass* 7 (2): 139–151.

Linnebo, Ø. 2018. *Thin Objects*. Oxford and New York: Oxford University Press.

Picardi, E. 2012. Construction and Abstraction. In G. Löffladt (Hrsg.), *Mathematik - Logik - Philosophie. Ideen und ihre historischen Wechselwirkungen*, 305–320. Frankfurt am Main: Wissenschatlicher Verlag Harri Deutsch.

Picardi, E. 2016. Michael Dummett's Interpretation of Frege's Context Principle. Some Reflections. In *Justification, Understanding, Truth, and Reality*, ed. M. Frauchinger, Berlin: De Gruyter.

Rayo, A. 2013. *The Construction of Logical Space*. Oxford and New York: Oxford University Press.

Rayo, A. 2014. Nominalism, Trivialism, Logicism. *Philosophia Mathematica* 23 (1): 65–86.

Rosen, G. 1993. The Refutation of Nominalism(?). *Philosophical Topics* 21 (2): 149–186.

Rosen, G. 2010. Metaphysical Dependence: Grounding and Reduction. In *Modality: Metaphysics, Logic, and Epistemology*, eds. B. Hale and A. Hoffmann, 109–136. Oxford and New York: Oxford University Press.

Rosen, G. 2011. The Reality of Mathematical Objects. In *Meaning in Mathematics,* ed. J. Polkinghorne. Oxford and New York: Oxford University Press.

Shaffer, J. 2009. On What Grounds What. In *Metametaphysics*, ed. D. Manley, D. J. Chalmers, and R. Wasserman, 347–383. Oxford and New York: Oxford University Press.

Schwartzkopff, R. 2011. Numbers as Ontologically Dependent Objects. Hume's Principle Revisited. *Grazer Philosophische Studien* 82: 353–373.

Wright, C. 1983. *Frege's Conception of Number as Object*. Aberdeen: Aberdeen University Press.

Wright, C. 1988. Why Numbers Can Believably Be. *Revue Internationale de Philosophie* 42: 425–473.

Wright, C. 1990. Field and Fregean Platonism. In *Physicalism in Mathematics*, ed. A. Irvine, 73–93. Dordrecht: Kluwer Academic Publishers.

Yablo, S. 2001. Go Figure: A Path Through Fictionalism. *Midwest Studies in Philosophy* XXV: 72–102.

Yablo, S. 2005. The Myth of the Seventh. In *Fictionalism in Metaphysics*, ed. M.E. Kalderon. Oxford: Oxford University Press.

Zalta, Edward N. 2017. Frege's Theorem and Foundations for Arithmetic. In *The Stanford Encyclopedia of Philosophy*, ed. N. Zalta Edward, Summer ed, https://plato.stanford.edu/archives/sum2017/entries/frege-theorem/.

A Context Principle for the Twenty-First Century

Fabrizio Cariani

For several generations of students at the University of Bologna, Eva Picardi's course in the philosophy of language has been a unique portal to a difficult subject matter and an exciting philosophical style. Her lectures on Frege, Russell and Wittgenstein were a treasure trove of bold, and yet seriously scrutinized, ideas about language and how it manages to connect up with reality. As one of these students, I went on to write an undergraduate thesis on Frege's context principle under her attentive and inspiring supervision. This was a fortunate choice, as she was about to devote a good chunk of her subsequent research to the context principle itself (Picardi 2002, 2009, 2016). In the years since, my interests within philosophy of language have shifted somewhat, but I want to use this opportunity to honor Eva's career and life by resuming that conversation.

The context principle (CP) is one of two core pieces of Frege's methodological legacy. It is often glossed as the requirement that one should never ask for the meaning of a word in isolation, but "only in the context of a proposition". It makes a (somewhat uneasy) pair with another

F. Cariani (✉)
Northwestern University, Evanston, IL, USA
e-mail: f-cariani@northwestern.edu

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_8

methodological injunction associated with Frege: the principle of compositionality, according to which the meaning of a sentence is composed of the meaning of its constituent expressions.

There are many important problems concerning the proper interpretation and philosophical significance of these principles, and we will get to some of them in order. But the one that got me thinking about *CP* after all these years is rather frivolous: there is a striking gap between the prime importance that Frege scholars ascribe to *CP*[1] and the relatively marginal role it appears to occupy in contemporary philosophy of language and natural language semantics. This is not because contemporary research paradigms have turned their back on Frege. On the contrary, he remains hailed as one the major pioneers of the field. Yet, very few of the main research programs in philosophy of language and semantics are explicitly premised on *CP*. This is in stark asymmetry with the principle of compositionality, which survives as a fundamental tenet in the model-theoretic tradition in semantics (see Szabó 2013 for an overview).[2] To put the point in a somewhat tongue-in-cheek way: I have seen scores of papers rejected from journals and conferences because of violations of compositionality, but I have never seen anyone complain about violations of the context principle.

I can see the sharpshooters lining up to shoot down these claims, so let me qualify them a bit. It is certainly not right to say that no one cares about the context principle. It has been openly embraced by inferentialists (Brandom 1994, sections 7.1–7.2; 2002, chapter 8) and by so-called radical contextualists (Travis 2009). However, it is not clear that these appeals are faithful to the context principle as Frege understood it. Picardi herself (2009, sections 1 and 2) makes a compelling case that Travis's version of the context principle depends on a richer interpretation of 'context' than what Frege had in mind. Serious doubts have also been raised about the accuracy of Brandom's interpretation (Kleemeier and Weidemann 2008). Even setting these interpretational matters aside, there are two points that are worth emphasizing:

(i) The model theoretic tradition in semantics has largely ignored *CP*. The principle is not mentioned in the dominant textbook in this tradition (Heim and Kratzer 1998). And while it does put in a surprising cameo in another important textbook (Chierchia and McConnell-Ginet 1990, section 3.1), it is under an interpretation

which is considerably weaker than what Frege seems to have had in mind.

(ii) Almost no one, regardless of philosophical preferences, ever invokes *CP* to argue *against* other views (it is used by those who use it to elaborate their own positions).

This essay proposes an interpretation of the context principle that aims to be as faithful as possible to Frege's writings (spoiler: I don't think we can be entirely faithful) while at the same time being of some value for the modern philosopher of language. My interpretation also aims to explain why the context principle is best used defensively (to articulate one's own views) rather than offensively (to criticize other views). Moreover, I will attempt to do so without relying very much on the role that *CP* plays in Frege's argument for mathematical Platonism.[3] I do not, of course, deny that *CP* is part of an argument for mathematical Platonism. Nonetheless, nothing in *Foundations* suggests that *CP* is *only* of significance for Platonists. If the principle is to be a valuable piece of the philosopher of language's toolkit, it should have purpose and significance when removed from this metaphysical application. If anything, we ought to generalize to the worst case and ask: can the principle be formulated so as to be compelling even for mild nominalists?

This is the plan: Section 1 lays down basic background concerning *CP*. Section 2 advances some constraints on an adequate interpretation of it. In particular, it follows several other interpreters (starting with Dummett 1956 and ending with Picardi 2016) in arguing that the slogan "Words only have meaning in the context of a proposition" cannot, literally speaking, be quite right. It will flow out of this argument that the principle must apply in the first instance to *explanations* of meaning—a conclusion which much of the literature already embraces. Section 3 entertains and rejects a recent suggestion to the effect that the context principle concerns *metasemantic* explanations (metasemantic interpretations of *CP* have recently been discussed by Stainton 2006 and Linnebo 2009). Section 4 defends an alternative: I argue that *CP* concerns what I will call *licensing* explanations. These are explanations that are required if a theorist is to accept a definition of a term as successful. In Section 5, I use the interpretation of the previous section to reassess some delicate questions about the scope of the context principle. In the concluding section, I will tie up some loose ends.

1 A Brief History of the Context Principle

Here is a familiar story: at four points in The Foundations of Arithmetic (1884, henceforth *Foundations*), Frege invokes a mantra connecting word meaning with "the context of a proposition". (I assume, with the rest of the literature, that Frege means 'sentence', and not what contemporary philosophers of language mean by 'proposition'.) The exact formulations of the mantra vary a bit[4]:

> *Introduction, p. X*: "ask for the meaning of a word in the context of a proposition, not in isolation; […] If one ignores [this principle], one is almost forced to take as the meanings of words mental pictures or acts of the individual mind".
>
> §60 "[…] any word for which we can find no corresponding mental picture appears to have no content. But we ought always to keep before our eyes a complete proposition. Only in the context of a proposition have the words really a meaning (*Bedeutung*)".
>
> §62 "How, then, are numbers to be given to us, if we cannot have any ideas or intuitions of them? Since it is only in the context of a proposition that words have any meaning, our problem becomes this: To define the sense of a proposition in which a number word occurs".
>
> §106 "We proposed the fundamental principle that the meaning of a word is not to be explained in isolation, but only in the context of a proposition: only by adhering to this can we, as I believe, avoid a physical view of number without slipping into a psychological view of it".

There is an important contrast between the passage from §106, which is explicitly about explanations of meaning, and the passages in §60 and §62, which are ostensibly about when words *have* meanings. (The passage from the introduction seems to be neutral, since "ask for the meaning…" could be understood either way).

Despite these differences in formulation, with the exception of §62, Frege is single-minded about what *CP* is supposed to do: those who ignore it fall into the murderous embrace of psychologism. They end up taking the meanings of number words to be mental pictures or ideas (call this the *defensive* use of *CP*). The passage from §62 departs from this talking point to pursue a more positive thought: numbers can be "given to us" if we fix the meaning of certain sentences involving numerals.

There is no doubt that any adequate interpretation of *CP* needs to make sense of both these roles it plays in the dialectic of *Foundations*. This is harder to do in the case of the positive use, because we need to simultaneously interpret the principle itself and gain an understanding of what it is for number to be "given to us". For this reason, I generally prioritize the defensive use in shaping my interpretation.

When Frege wrote *Foundations*, he had not yet drawn his celebrated distinction between sense (*Sinn*) and reference (*Bedeutung*)—its official formulation must, of course, wait until his (1892). Naturally, this raises the question whether by using 'Bedeutung' in the 1884 mantra Frege means to state a principle that concerns (i) the technical semantic notion of reference; (ii) the technical semantic notion of sense; or (iii) the non-technical and undifferentiated notion of meaning. Unfortunately, the essays from the 1890s are of little help in resolving this issue because, by the nineties, Frege stops explicitly appealing to the context principle.[5]

Faced with this difficulty, interpreters have converged on a reasonable approach: take *CP* to involve the undifferentiated notion of meaning, but also ask whether the internal logic of Frege's argument suggests an interpretation in terms of sense or one in terms of reference.

Frege might have quit invoking the mantra cold turkey, but his followers in the analytic tradition had a different idea. Wittgenstein endorses the mantra in *Tractatus Logico-Philosophicus* (1921) and in the *Philosophical Investigations* (1946). Quine, too, reached for an interpretation of *CP* in *Two Dogmas of Empiricism* (1951, p. 39) and *Epistemology Naturalized* (1969, p. 72). Among the interpreters, Dummett is most famous for having emphasized the importance of the doctrine within Frege's thinking (this was, remarkably, a four-decades long engagement, see Dummett 1956, 1973, 1981, chapter 19; 1991, chapters 16–17; 1995).

Put this all together and you get a cluster of interrelated questions:

(Q1) how exactly does *CP* figure in the *Foundations*' critique of psychologism.
(Q2) did Frege give up *CP* after *Foundations*?
(Q3) how does *CP* interact with the *Sinn/Bedeutung* distinction?
(Q4) Is *CP* compatible with the principle of compositionality?
(Q5) Did *CP* play a similar role in Wittgenstein's philosophy as it did in Frege's philosophy?
(Q6) Is *CP* a separable component of Frege's philosophy of language?
(Q7) Is *CP* something that contemporary theorists should obey?

Some of these questions have simple, clear and convincing answers. As I noted, for instance, Picardi (2009) argues convincingly that, due to the ontology of propositions in the *Tractatus*, the early Wittgenstein's invocation of the context principle could not possibly be in the same spirit as Frege's. Moreover, as far as the *later* Wittgenstein is concerned, the notion of "context" he employed when he repeated the mantra seems to have been altogether different from Frege's. In particular, Wittgenstein's contexts include *extra-linguistic context*. There is little chance that this might have been on Frege's mind in the *Foundations*. Most of the other questions are still open despite extensive debate. In the rest of this essay, I re-elaborate some familiar answers to (Q1), (Q3), (Q4) and sketch some answers to (Q6) and (Q7).

Before proceeding, let me make some terminological stipulations. First, as is well known, Frege assigns the same semantic profile to proper names ('Michelle Obama') and definite descriptions ('the most famous alumna of Whitney Young High School'). I use the phrase *singular terms* to cover both. Second: the context principle concerns the meanings of sub-sentential expressions, a category that includes both individual words and phrases like definite descriptions. Despite that, in the following, I use the word *word* as a substitute for 'sub-sentential expression'. This is in part to improve readability and in part because that's what Frege uses in the passages I quoted above. Finally, *Foundations* discusses two kinds of singular terms: numerals (e.g. '2') and descriptions formed by means of the cardinality operator ('the number of Jupiter's moons'). I use the phrase *numerical terms* to cover both of these.

2 How Not to Interpret the Context Principle

One of the main reasons why there is an interpretive problem about *CP* is that the literal interpretation of the principle cannot be quite right. Taken literally, the mantra "Only in the context of a proposition does a word have a meaning" seems to mean something like this:

> **Literal:** meaning is something that word it has only as part of full sentential contexts. Take it out of sentential contexts and it does not have a meaning.

As noted above, you might get specialized versions of this thesis by replacing the undifferentiated notion of "meaning" with technical notions like sense and reference.

Here is one possible analogy for this sort of thought: which direction is this arrow pointing to?

→

This question only makes sense in a broader context: for example, it could be pointing North, or towards the center of the earth, towards the right side of the page, or towards Alpha Centauri. There is no sense to isolating that object, the arrow-token, and trying to figure out where it is pointing. Similarly, there is no sense to isolating a G# note and asking whether it's a dominant. Only in the context of a scale, or a chord, can a note be a dominant. What is common between these cases is that these the relevant properties seem to have some kind of implicit or explicit relationality.[6] Maybe word meaning is like that.

I don't think this can quite be the point of Frege's version of the context principle. For one thing, this seems obviously false (Linnebo 2009). Consider proper names. There is no theory of proper names on which it's impossible to answer the question "what is the meaning of the name 'Michelle Obama'?". If you think that the meaning is the referent, then that's it. But suppose, you do not. Say that you believe instead that the meaning of the name has two elements, a referent (the woman herself) and a sense (a mode of presentation of the referent). Then that referent and that sense together are the meaning of 'Michelle Obama'. From the contemporary, non-interpretive point of view, the literal interpretation is also incomprehensible: as theorists of meaning and even as ordinary users of the language we can sometimes ask for the meanings of words: there is no reason to suppose that Frege's texts point to a relevantly different practice.

The second problem is that the literal interpretation makes no sense of the defensive use of *CP* in the *Foundations* (or, for that matter, of the positive use). Recall that Frege's negative aim is that we need not expect to explain the meaning of number words by associating them with perceptually given referents. It is entirely irrelevant to that target to assume that a word *e* lacks meaning when taken on its own.

As Linnebo (2009) puts it, when it comes to the context principle, we must forget the literal reading of the mantra and instead look primarily at how Frege uses it. Furthermore, concerns about reaching for non-literal interpretations ought to be assuaged by noticing (as we did earlier) that the literal interpretation is only encouraged by the formulations in §60 and 62. The other two formulations, in the *Introduction* and in the summary (§106), are compatible with the denial of the literal interpretation.

At the other end of the spectrum, there is a plausible but weak interpretation of *CP* that cannot quite do the work that *CP* is supposed to do. It is often said that *CP* states the primacy, or priority, of sentential meaning over word meaning. Here are a couple of examples drawn from two very different places:

> The Context Principle, as one relating to sense, amounts to the conceptual priority of thoughts over their constituents: the constituents can be grasped only as potential constituents of complete thoughts. (Dummett 1991, p. 184).

> The Context Principle […] seems to be that (i) the notion of a sentence having a meaning (which Frege identifies with the claim it is used to assert or express) is explanatorily primary, while (ii) what it is for a word or phrase to have a meaning is to be explained in terms of what it is it contributes to the meanings of sentences containing it. (Soames 2014, p. 47)

Unfortunately, that claim is about as equivocal as the word "primacy" is underspecified. Perhaps, the relevant sense of primacy is this:

> **Weak context principle (*WCP*)**: the meaning of a word *e* is *e*'s contribution to sentential meaning.

For an analogous case, we might say that the tactical objectives of a striker in soccer are that striker's contribution to the team's tactical objectives. Once again, this can be declined into a thesis about reference (the reference of a word is its contribution to the reference of the sentence) and into a thesis about sense (the sense of a word is its contribution to the senses of the sentences it contains).

WCP is far from vacuous: if sentential meanings are truth-conditions, then the meaning of a predicate is its contribution to the truth-conditions of a sentence. If the sense of a sentence is a structured proposition, then the sense-variant of *WCP* requires that the sense of a name be what the name contributes to that structured proposition. Still, the idea is pretty plausible and arguably satisfied by any modern theoretical account of linguistic meaning.

I do not doubt that *WCP* is part of what is involved in the context principle. I doubt, however, that it can capture the whole content of *CP*. As in the case of the literal interpretation, it is not clear how this

formulation accounts for the negative application of *CP* in *Foundations*. Suppose you did *not* think that a word's meaning is to be identified with its contribution to sentence meaning. Why would that force you to identify the meaning of the numeral '3' with a mental picture? No explanation of this is forthcoming because the supposition we are operating under is so open-ended that it provides no guidance. Part of the problem here is that *WCP* is a constraint that simultaneously operates on word meaning and on sentence meaning. Nothing specifically follows from it about word meaning, unless we make assumptions about sentence meaning. The context principle must have content that goes beyond that of *WCP*.

As for what this extra content is, I think the basic idea is well put in Linnebo (2009): the Context Principle in *Foundations* concerns in the first instance *explanations of meaning*. It's not that you can't ask the question: "What is the meaning of a word *e*?". It's that you can't expect an explanatory answer to that question to be wholly independent of the semantic properties of sentences.

3 Metasemantic Explanation

But what kind of explanation is at stake? And what exactly needs explaining? In this section, I discuss the hypothesis that Frege is concerned with *metasemantic* explanation.

Metasemantics, broadly understood, is the study of the facts in virtue of which semantic properties hold. It is the job of a metasemantic theory, for instance, to characterize the facts in virtue of which the word 'koala' has the semantic value that it does. The distinction between semantic and metasemantic inquiry is nicely illustrated by the case of non-verbal signals. Contrast these two dialogues:

> **Semantic inquiry**. *Q*: what is the meaning of a red octagonal street sign with the word 'STOP' written across it? *A*: cars approaching the sign are required to stop before proceeding through the intersection.

> **Metasemantic inquiry**. *Q*: why does that sign have that meaning? *A:* Because the traffic authority established (and enforce) this conventional association.

Although the label is relatively recent, many classical theories in philosophy of language are metasemantic. This is the case for the sophisticated

conventionalism defended by Lewis (1975), according to which the conventions that underlie linguistic meaning are regularities of behavior embedded within a particular structure of mutual expectations.[7] Similarly, Stalnaker (1998) argues that Kripke's causal theory of reference is best viewed as a metasemantic hypothesis. The same is true of Lewis's doctrine of *reference magnetism* (Lewis 1983)—according to which part of what makes an assignment of meanings to predicates correct is the degree of naturalness of the properties they are to denote.

According to the metasemantic interpretation, the context principle constrains explanations of why words have the meanings they have.[8]

> **Metasemantic Context Principle (MCP)**: word *e* has meaning *m* in part because a range of sentences involving *e* are meaningful.

According to this thesis, word meaning is grounded in sentence meaning. Numerical terms (like 'the number of registered vehicles in Chicago') mean what they do because a range of sentences involving them are meaningful. In particular, these sentences must involve all the identity statements in which such terms can figure. The metasemantic context principle need not be one's entire metasemantic theory: there might be further facts in virtue of which those core sentences are endowed with meaning. Moreover, presumably, at some point, usage among the members of a linguistic community should show up in the metasemantics. That would constitute an additional layer of the theory.

If *MCP* generalizes beyond numerical terms, say to cover names of concrete objects like 'Michelle Obama', then it claims that the facts in virtue of which 'Michelle Obama' means *m* include the fact certain sentences have meanings. A further question still is whether the metasemantic context principle should extend to other syntactic categories.

Though far from obvious, and clearly in need of defense, *MCP* is an interesting thesis for the modern philosopher of language to evaluate. It also has some benefits as an interpretation of Frege: under this interpretation, the context principle is not really in tension with the principle of compositionality. Given compositionality, the meanings of complex expressions must depend on—and perhaps be composed of—the meanings of their constituents. But that's compatible with the idea that the facts that ground word meanings are partly determined by sentence meanings. The dependence that is claimed in compositionality and the reverse dependence that is claimed by *MCP* operate on different levels.[9]

MCP is a very strong claim about the nature of semantic properties. It would be extremely surprising if Frege thought that such a claim could be stipulated without any kind of justification. I suspect that this is a reason to doubt that *MCP* could possibly be what Frege had in mind. Of course, the fact that the context principle is presented without justification is an embarrassment for virtually every interpretation. Still, we should be wary of any proposal that requires *CP* to be much stronger than the argument demands, and, unfortunately, *MCP* is much too strong for the demands of the argument. The argumentative goal of §§58–60 of *Foundations* is to establish the thesis that it is possible to refer to objects of which we cannot form any sensible images. If *MCP* were right, it would deliver an explanatory account of how numerical terms in fact *do* refer to these objects. In other words, at a point where the argument merely demands a proof of concept of sorts, Frege would fire back with an unsupported and sweepingly general metasemantic thesis about how they in fact work.

But what other kind of explanation of meaning might we seek? Let me try to approach my view by thinking through a concrete example. Suppose I gather the people of the Earth and declare that 'wowee' is to be a new singular term. I claim that it is to denote the largest star in the universe. Imagine that I have the authority to get a linguistic practice going this way. After my declaration people start using 'wowee' just in the way I said. Imagine further that it is common knowledge between me and my audience that I do not have a particular star in mind which I believe to be the largest in the universe.

It seems to me that this scenario largely settles the metasemantic questions. There might or might not be a unique largest star. But if there is, I am successful in referring to it—whether I know that such a star exists or not. In this case, then, we have a semantic fact:

(F) 'wowee' refers to the largest star

We can ask the metasemantic question: why does (F) hold? And the story stipulates a simple answer: somehow, I have the power to get a linguistic practice going which helps link up that word with things; I have used this power to declare that 'wowee' is to refer to this celestial body.

But it is possible to press another question—one that parallels the question Frege is asking in §§58–60: how might one know whether my declaration successfully established that 'wowee' has a referent? That

requires a different sort of account—in particular it requires an argument to the effect that there is exactly one star with maximal size. This account is not itself part of metasemantics, since it need not have anything to do with the facts in virtue of which 'wowee' refers to the largest star. Suppose that my friend Alexis knows that there is a unique largest star. Alexis is in a position to vouch that my introduction of 'wowee' was successful. But there are many ways for Alexis to have obtained the relevant knowledge. Maybe she derived it from the laws of physics, maybe she was able to measure the sizes of all the stars in the universe, or maybe a very knowledgeable being whispered it in her ear. None of these justifications belong to the metasemantic story. There is typically *one* correct metasemantic story about why words mean what they do, but there are many ways we might gain the knowledge that the metasemantic fact holds.

4 Licensing Explanations

My view is that Frege believed that successful definition of numerical terms required the sort of guarantee that Alexis is able provide in the case of 'wowee'. In fact, this sort requirement shows up explicitly in a famous a passage from *Über Sinn und Bedeutung*:

> A logically perfect language (*Begriffsschrift*) should satisfy the conditions that every expression grammatically well constructed as a proper name out of previously introduced signs shall in fact designate an object, and that no new sign shall be introduced as a proper name without being secured a reference (*Bedeutung*). In logic, one must be wary of ambiguity of expressions as a source of mistakes. I regard as no less pertinent against apparent proper names that have no reference (*Bedeutung*). The history of mathematics supplies errors which have arisen in this way. This lends itself to demagogic abuse as easily as ambiguity […]. 'The will of the people' can serve as an example; for it is easy to establish that there is […] no generally accepted reference (*Bedeutung*) for this expression. (p. 41, slightly modified from the Geach translation)

Of course, in natural language we don't often go about introducing new terms as one would do in a formal language. But the project of *Foundations* is neither to describe a human linguistic practice, nor to specify the semantics for a formal language. It is to provide an analytic

derivation of the truths of arithmetic based, among other things, on definitions for numerical terms. To be confident that this project is in order, we need a justification similar to the one Alexis provides in the fictional example of 'wowee' and to the one Frege claims we lack in the case of 'the will of the people'. To state the requirement a bit more precisely, this is what we need (here "we" means "mathematical philosophers engaged with the Fregean definition project").

> **Definitional licensing**: we are licensed to claim that a definition of singular term *t* has provided it with meaning only if we have a guarantee that the object that *t* purports to refer to exists and is unique.

This kind of check—for the existence of referents of defined singular terms—is both standard mathematical practice and implicit in Frege's procedure in *Foundations*.

So far, I have attributed to Frege the view that, to accept a definition as successful, we must provide evidence that its referent exists and is unique. But what does the context principle have to do with all of this? According to my interpretation, the main role of the context principle in the *Foundations* is to highlight an easy-to-miss way of satisfying this requirement. Specifically, in §§58–60 Frege is pushing back against this argument:

(P1) A definition of numerical term *t* can be recognized as being correct only if we have a guarantee that the object that *t* purports to refer to exists and is unique.
(P2) We have such a guarantee only if numbers are sensible objects.
(P3) But numbers are not sensible objects.
(C) So, a definition of *t* cannot be recognized as correct.

Frege endorses (P1) and (P3), but rejects (C). The context principle is invoked in order to justify rejection (P2) and support the thesis that, once you have secured truth-conditions for a broad enough class of sentences—crucially a class containing all the identities between numerical singular terms—the definitional licensing requirement is satisfied.

Summing up, the import of the context principle can be characterized as follows:

> **Licensing Context Principle (*LCP*)**: the requirement of definitional licensing concerning a singular term *t* is satisfied if we have a guarantee that all of the identity statements involving *t* are meaningful.

Once this perspective is adopted, some of the puzzles surrounding *CP* get immediate answers. For example, we asked at the outset whether the principle applies to the notion of reference. Given the interpretation I am advancing, this is inevitable. This is because among the essential constraints that the definition of term *t* must satisfy is an argument that it, in some sense, it refers.

I add the qualification "in some sense", because if this is a requirement that can be cleared by the sort of strategy suggested by the context principle (i.e. by setting the meanings of certain sentences), this is a notion of reference that even a moderate kind of nominalist could live with. My sympathies here are strongly with Dummett's (1991, chapter 15) contention, echoed by Picardi herself (2016), that the theoretical framework of *Foundations* lacks the resources to distinguish between a robust realist construal of the reference relation and a more moderate one.[10]

Among the questions that we are in a position to address is also the vexed matter of whether the context principle is compatible with compositionality. The answer is that it is compatible, for the broadly Dummettian reason that the two principles claim explanatory priority in two different senses. Compositionality is about how semantic values depend on each other, while *CP* states that *sometimes* we go about justifying the meaningfulness of a definition or of a linguistic practice by pointing to the semantic properties of certain sentences. Similarly, we can explain how within the context of *Foundations* Frege can at the same time maintain that he has substantively used the context principle and yet provide an explicit definition of numerical terms. After all, under the present interpretation is simply not a constraint on the form of the definition.

5 The Limited Scope of the Context Principle

Much of this aligns with some of Dummett's influential views about the role of the context principle. But on one issue, I think the above discussion requires a revision of Dummett's position. Once we adopt the view that the context principle should be mainly interpreted as in *LCP*,

two questions arise concerning its scope. The first question is: does *LCP* apply to all singular terms? Or could we instead have a tempered version of the principle that only applies to singular terms denoting mathematical and other abstract objects? The other important question is whether *LCP* applies to expressions across all syntactic categories (e.g., does it apply to predicates, function symbols, quantifiers)?

With regards to the first question, we need to disambiguate between two possible ways of hearing the word "apply". Again an analogy is helpful. Imagine a fictional country, the Shire, with these laws:

> **Norm 1**. Voting in the Shire requires citizenship in the Shire
>
> **Norm 2**. One has citizenship in the Shire if one has resided in the Shire continuously for five years.

The logical structure of these norms is the same as the one I have sketched for the context principle: Norm 1 lays down a requirement and Norm 2 articulates one way of satisfying that requirement. Now, consider Frodo who has resided in the Shire for less than five years: does Norm 2 apply to him? In one sense, the answer is "no": Frodo does not satisfy the condition stipulated in the antecedent of Norms 2. If he meets the requirement for voting in Norm 1's requirement, it is not because he satisfies the condition specified in Norm 2. In another sense, the answer is "yes": Norm 2 applies to Frodo just as much as it applies to anyone else who lives in the shire. You don't need to satisfy the antecedent of the norm for it to apply to you.

The case of the context principle is structurally analogous: *LCP* characterizes a way for us to be satisfied that a certain expression is meaningful. As such, it applies with equal force to every term. If we wanted to introduce the name 'Michelle Obama' in our linguistic practice we can acquire the relevant license by following the path indicated by *LCP*. But that does not mean that there are no other ways for us to acquire that license. When we look at the variety of singular terms in our language we need not expect that the Licensing requirement is satisfied in the same way for every term.

One unexpected implication of my reading of *Foundations* is that it is consistent to think that the name 'Michelle Obama' is licensed by direct contact with the object. This puts me at odds with mainstream views about the interpretation of the context principle. Here for instance, is Wright:

> The really fundamental aspect of Frege's notion of object and concept is that they are notions whose proper explanation proceeds *through* linguistic notions. (Wright 1983, p. 13)

This is an intriguing picture, but it comes with the huge interpretive cost that Frege never tries to the defend it, as Wright himself acknowledges (1983, p. 15).

My interpretation is also likely at odds with one of the conclusions of Dummett's interpretation:

> The realist interpretation could be jettisoned without abandoning the context principle itself, but only if that principle, as here understood, can be shown to be coherent; and this remains in grave doubt. And yet it is hard to see how it can be abandoned, so strong is the motivation for it. The alternative is an apprehension of objects, including abstract objects, underlying, but anterior to, an understanding of reference to them or, indeed, a grasp of thought about; and this is a form of realism too coarse to be entertained. (Dummett 1995, p. 19)

Strictly speaking, I agree that the context principle is meant to show that "direct apprehension" cannot be the *only* means of securing meaning for singular terms. But a stronger claim is implied here: that there is something wrong with a mixed picture on which the licensing requirement is sometimes satisfied via the context principle, sometimes via direct apprehension. I do not think that anything in Frege's discussion justifies this.

Something similar should be said about the question whether the context principle applies across syntactic categories. When the semantic theorist claims, for instance, that the semantic value of "smokes" is a function from individuals to truth-value that maps smokers to the true and non-smokers to the false, we need some kind of assurance that such a function exists. In principle, we could provide this assurance through *CP*-like reasoning. But in practice, we do so by relying on basic assumptions about existence of individuals and truth-values as well as function-existence assumptions.

6 Conclusions

To conclude, I want to connect this discussion back to the issues that prompted me to rethink about the context principle and to the main outstanding questions from my earlier list:

(Q6) Is *CP* a separable component of Frege's philosophy of language?
(Q7) Is *CP* something that contemporary theorists should obey?

An affirmative answer to (Q6) is possible, but something must be added to the story I just told. Because I limited my interpretation to *Foundations*, questions about natural language have not really been within our theoretical sights. When we theorize about natural language, the constraints that guided Frege's thoughts on definitions need not apply. Still, you might ask a parallel question to the one that is behind the requirement of definitional licensing: how do I know that some term *t* of my language refers? And one possible answer might be that this can be known by being able to fix the truth-conditions of certain sentences involving the term—without actually having any causal contact with the referent of the term. As with all these views, it is impossible, and not entirely intellectually honest, to say that one is for, or against, them, in advance of actually spelling out their detail.

As for (Q7), I think that the answer is that there is no such thing as obeying the context principle. If my interpretation is correct, the principle is meant to point to a way of satisfying certain requirements—it is not itself a requirement. Still, one may ask if this is a path to satisfying the requirement that any contemporary theories exploit. The answer I provide to this question is the same that would be given by a proponent of the metasemantic interpretation: one cannot simply look at one's favorite semantic model and extrapolate whether the context principle is satisfied. The difference is that the metasemantic interpretations would say that the context principle might show up when we tell the story about what is it in virtue of which singular terms denote what they denote. On my preferred story, the context principle can help when we ask how we are entitled to the belief that various terms of our language have denotations and, more generally, meanings. This is typically going to be particularly pressing for names of abstract objects, if there are any, for in that case, the causal route is blocked.

Acknowledgements I thank Paolo Santorio and Rob Stainton for conversations about this material, as well as Annalisa Coliva, Paolo Leonardi and Sebastiano Moruzzi for editing this volume and inviting me to contribute. My biggest debt is to Eva herself: I owe her for the mentorship, the conversations, for helping me track down some of her writings on this subject, but most of all for inspiring me to become, like her, a quietly irreverent philosopher.

Notes

1. As we will see, Dummett's engagement with the context principle spanned his entire career, but see Pelletier (2001) for a remarkable catalog of authors.
2. Frege scholars draw a distinction between the *principle of compositionality* (that the meanings of sentences are composed of the meanings of their parts) and the *principle of functionality* (the meanings of sentences are functionally determined by the meanings of their parts). When theorists in the model-theoretic semantics traditions talk about 'compositionality' they usually refer to something closer to functionality, since compositionality properly understood requires meanings to have mereological structure.
3. Discussions of the context principle and its relationship to definition by abstraction have been central to the Neo-Fregean program. Key references include Wright (1983), Hale (1987), Boolos (1990), Hale and Wright (2001). This debate goes quite beyond the relatively limited set of questions I intend to ask about the context principle here.
4. English translations are mostly Austin's with some touch-ups for clarity and accuracy.
5. That is to say: he stops using the mantra. One of the main fault lines in the secondary literature concerns whether there are passages that could be viewed as later occurrences of *CP*. Resnik (1967) claims that Frege gave up the context principle; by contrast, Dummett might have contributed the most to an argument that a version of the doctrine survives in his later works. The most plausible view of the matter seems to be this: there is no explicit reliance on the principle but that doesn't mean that he wasn't relying on the same kind of insight (Dummett 1981, chapter 19). Still, this means that the later work is of little help in interpreting the principle. This is because, from this perspective, we must, first, interpret the principle as it occurs in *Foundations*, and then evaluate whether there are instances of similar reasoning in later works.
6. This relationality seems different from the concept of relational properties that is prevalent in metaphysics. On that usage, a property P is relational just in case an object's having P depends essentially on its standing in certain relations to other objects. For example, *being a husband* is such a property. The examples in the text, instead, are cases of properties whose applicability to objects presupposes certain other facts.
7. Also metasemantic is the naïve conventionalism mocked by Quine in the introduction to Lewis's "Convention": "When I was a child I pictured out language as settled and passed down by a board of syndics, seated in grave convention along a table in the style of Rembrandt".

8. This interpretation of *CP* is defended by Linnebo (2009) and, before that, it was critically explored by Stainton (2006).
9. See Linnebo (2009, §6.2). Linnebo imports this equivocation account of the conflict into the metasemantic picture from an analogous diagnosis of equivocation by Dummett (1973, pp. 3–7) and Dummett (1981, p. 547).
10. Picardi also draws attention to an important discussion in Carnap's *Aufbau*. Carnap (1928, section 40) was happy to take contextual definitions on board, provided that numerical terms were understood as denoting "quasi-objects". "Carnap's (1928) interpretation of Frege's conception of classes captures important features of [*Foundations*], and indirectly supports Dummett's diagnosis that only a thin conception of reference is appropriate to [*Foundations*]. I find it difficult to decide which way we should go: [*Foundations*] seems hospitable both to Carnap's and to Wright's interpretations, for it is semantically underdetermined" (Picardi 2016, p. 54).

Bibliography

Boolos, George. 1990. The Standard of Equality of Numbers. In *Meaning and Method: Essays in Honor of Hilary Putnam*, ed. G. Boolos. Cambridge, MA: Harvard University Press. Reprinted in George Boolos, *Logic, Logic and Logic*, Harvard University Press, 1998.

Brandom, Robert. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.

Brandom, Robert. 2002. *Tales of the Mighty Dead*. Cambridge, MA: Harvard University Press.

Carnap, Rudolf. 1928. *Der Logische Aufbau der Welt*. Felix Meiner Verlag [English translation, *The Logical Structure of the World; Pseudoproblems in Philosophy*]. Berkeley: University of California Press, 1967.

Chierchia, Gennaro, and Sally McConnell-Ginet. 1990. *Meaning and Grammar*, 2nd ed. Cambridge, MA: MIT Press, 2000.

Dummett, Michael. 1956. Nominalism. *The Philosophical Review* 65 (4): 491–505.

Dummett, Michael. 1973. *Frege: Philosophy of Language*, 2nd ed. Cambridge, MA: Harvard University Press.

Dummett, Michael. 1981. *The Interpretation of Frege's Philosophy*. Cambridge, MA: Harvard University Press.

Dummett, Michael. 1991. *Frege: Philosophy of Mathematics*. Cambridge, MA: Harvard University Press.

Dummett, Michael. 1995. The Context Principle: Centre of Frege's Philosophy. In *Logik und Mathematik: Frege-Kolloquium Jena 1993*, ed. I. Max and W. Stelzner. Berlin: de Gruyter.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik* [English translation, Translated by J.L. Austin, *Foundations of Arithmetic*]. Oxford: Blackwell, 1953.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. In *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50. Translated as 'On Sense and Reference' by M. Black in Translations from the Philosophical Writings of Gottlob Frege, P. Geach and M. Black (eds. and trans.), 3rd ed. Oxford: Blackwell, 1980.

Hale, Bob. 1987. *Abstract Objects*. Oxford: Blackwell.

Hale, Bob, and Crispin Wright. 2001. *Reason's Proper Study*. Oxford: Clarendon.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Malden, MA: Blackwell.

Kleemeier, Ulrike, and Christian Weidemann. 2008. In *Robert Brandom: Analytic Pragmatist*, ed. Bernd Prien and David P. Schweikard. Ontos.

Lewis, David. 1975. Languages and Language. In *Minnesota Studies in the Philosophy of Science*, ed. K. Gunderson, 3–35. Minneapolis: University of Minnesota Press.

Lewis, David K. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61: 343–377.

Linnebo, Øystein. 2009. Frege's Context Principle and Reference to Natural Numbers. In *Logicism, Intuitionism, and Formalism: What Has Become of Them?* ed. S. Lindström, E. Palmgren, K. Segerberg, and V. Stoltenberg-Hansen, 47–68. The Netherlands: Springer.

Pelletier, Francis Jeffry. 2001. Did Frege Believe Frege's Principle? *Journal of Logic Language and Information* 10: 87–114.

Picardi, Eva. 2002. Il Principio del Contesto in Frege e Wittgenstein. In *La Svolta Contestuale*, ed. C. Penco, 1–23. Milano: McGrawHill.

Picardi, Eva. 2009. Wittgenstein and Frege on Proper Names and the Context Principle. In *Wittgenstein, Mind, Meaning and Metaphilosophy*, ed. P. Frascolla, D. Marconi, and A. Voltolini, 167–188. Basingstoke, UK: Palgrave Macmillan.

Picardi, Eva. 2016. Michael Dummett's Interpretation of Frege's Context Principle: Some Reflections. In *Justification, Understanding, Truth, and Reality: Themes from Dummett*, ed. M. Frauchinger. Berlin: De Gruyter.

Quine, W.V.O. 1951. Two Dogmas of Empiricism. In *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Resnik, Michael. 1967. The Context Principle in Frege's Philosophy. *Philosophy and Phenomenological Research* 27: 356–365.

Soames, Scott. 2014. *The Analytic Tradition in Philosophy, Volume 1: The Founding Giants*. Princeton University Press.

Stainton, Rob. 2005. The Context Principle. In *Encyclopedia of Language and Linguistics*, 2nd ed., vol. 3, ed. K. Brown, 108–115. Oxford: Elsevier.

Stainton, Rob. 2006. *Words and Thoughts.* Oxford and New York: Oxford University Press.

Stalnaker, Robert. 1998. Reference and Necessity. In *Blackwell Companion to the Philosophy of Language*, ed. B. Hale and C. Wright, 534–554. Oxford: Blackwell.

Szabó, Zoltán Gendler. 2013. Compositionality. *The Stanford Encyclopedia of Philosophy* (Fall 2013, Edition), ed. Edward N. Zalta. https://plato.stanford.edu/archives/fall2013/entries/compositionality/.

Travis, Charles. 2009. To Represent as so. In *Wittgenstein's Enduring Arguments*, ed. E. Zamuner and D.K. Levy, 4–29. London: Routledge.

Wright, Crispin. 1983. *Frege's Conception of Numbers as Objects.* Aberdeen University Press.

Slurs and Tone

Ernie Lepore and Matthew Stone

1 Introduction

Two claims that are hard to deny are that slurs can be offensive, and that not all uses of language are communicative. It's therefore perplexing why no one has considered the possibility that slurs might be offensive *not* because of what they communicate but rather because of interpretive effects their uses might exact. In what follows, we intend to argue just that, namely, that confrontations with slurs can set in motion a kind of

Professor Eva Picardi made invaluable contributions to the discussion of tone; in particular, we have been greatly influenced by her (2007). We are indebted to her for these contributions and in the case of one of us (EL) for her friendship dating back to May, 1985, including quite often a rich exchange of philosophical ideas. She was a joy to be with, as well as to talk philosophy with. Her untimely death is a great loss for all philosophers and even more so for her many devoted friends. Obviously, our contribution to this volume, based on, and partly extracted from (Lepore and Stone 2017), is dedicated to her.

E. Lepore (✉) · M. Stone
Rutgers University, New Brunswick, NJ, USA
e-mail: lepore@ruccs.rutgers.edu

M. Stone
e-mail: mdstone@cs.rutgers.edu

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_9

imaginative engagement that rouses objectionable psychological states. We believe this view has precedence in Frege's account of tone.

From early on, Frege recognized there is more to language than expressing thought and performing speech acts (Frege 1897, 139). For example, he noted a difference between 'dog' and 'cur'—expressions that, though the same in meaning, 'puts us…in mind' different associations. Anyone who uses 'cur' 'speaks pejoratively, but this is not part of the thought expressed' (Frege 1897, 140).[1] That is, 'a difference in tone need not, and often cannot, be spelled out by means of a specific description…' (Picardi 2007, 495).

Looking forward, our view is that tone is a catchall for interpretive effects that go beyond meaning, effects that are open-ended and even often non-propositional. This is not what Frege meant by tone, since one way of summarizing our view is that tone is *not* a coherent theoretical category: tones can have different sources and different consequences.[2] Nevertheless, there are reasons for *not* treating the pejorative tone of slurs as a remnant of linguistic meaning. In addition, our account complements, and amplifies on, the *Prohibitionist* view of slurs in Anderson and Lepore (2013a, b), who argue that the *only* distinctive status slurs share is their prohibition.

For critics of Prohibitionism, slurs are banned because something is wrong with them—not the other way around. Some critics complain that mere prohibition fails to distinguish slurs from other banned terms, e.g., from profanity (Camp 2013, 343). And even after narrowing the discussion to slurs, some critics still complain that Prohibitionism seems inadequate to account for their rhetorical power on their victims and witnesses (Camp 2013, 343). To say slurs are prohibited—and no more—leaves these matters unexplained. The obvious alternative has been that slurs are offensive because of what they mean. Accordingly, their semantics (or the pragmatics of their uses) prompts their ban, distinguishes their offense from profanity, and locates their uses on a scale of repugnance.

We believe Anderson and Lepore (2013a, b) have already established that, though meaning takes a variety of forms, none can accommodate the demeaning effect of slurs. We will go further, arguing that differences in Fregean *tone* among synonyms cannot be captured semantically (or pragmatically). Critics have missed this because they have failed to appreciate the range of options open for distinguishing the interpretive effects words elicit, even when identical in linguistic meaning. In what follows, we will take up this diversity.

2 The Variety of Sources of Tone

Consider the difference in uses between the words 'mother' and 'mommy'. In their basic senses, these words pick out the same individuals. There may be some linguistic differences corresponding to formality and informality of usage, but, as Jorhana Horn (2013) exclaims, 'there's a pure sweetness inherent in a small person looking up at you, calling you 'Mommy.' The word radiates innocence. It gives me the same feeling I get when I hold either of my daughter's little hands in my own. Every time G says it, I feel like she is putting a crown on my head–even when it's a whine…. 'Mommy!' is what a child calls out in the dark when they wake up from a nightmare, confident that the person who knows how to fix everything will come in and make the bad dreams go away.' These sentiments obviously go well beyond the demands of language. What Horn's comments make clear is that the difference between these two words is one of *perspective*. It is clear that they somehow *invite* audiences to think of a female parent in different ways. This is not a difference in *content*, and so, cannot be explained in terms of the nuanced articulation of linguistic content of contemporary semantic theory. But how do we turn these observations into a view? To this end, we first propose that difference in tone *alone* leaves synonymy, pure and simple.

Looking ahead, for us, slurs and their neutral counter-parts are identical in meaning. As far as language goes, what distinguishes slurs from neutral counterparts is only their prohibition. We could end the discussion here, but we want to elaborate on how it is possible for words to differ in tone and *yet* agree in meaning. When and where do these differences surface? If your habit is to locate interpretive differences in either semantics or pragmatics alone, then you might think our questions unanswerable. In fact, we believe the reverse is so. We turn to the variety of sources of tone.

One source of tone is a *difference in perspective*; e.g., think of the different perspectives that helps us to distinguish 'mother' from 'mommy'. Being a mommy, of course, indicates a certain biological etiology; or a certain role in child rearing; or even as the object of affection of an innocent child. And so on. Perspectives, so understood (Lepore and Stone 2015, following Camp 2006, 2008, 2009), are obviously *open-ended*, *non-propositional* psychological constructs; and as such they needn't encode any information about how the world is. All mothers and mommies are biologically connected to their children; all rear their children;

and all are the first object of affection for their children. Therefore, these different perspectives will not distinguish some mommies from some mothers—but these two words, just the same, can, and do, draw our attention to different aspects of motherhood, and influence how we think about what they apply to.

For another example, consider someone who says, 'Bruce Wayne is here.' Her utterance prompts different imaginings in an audience than someone who says, 'Batman is here.' Identical situations make both utterances true, but with the first, we can't help but imagine a suave American billionaire playboy, and with the second, a genius with physical prowess, martial arts abilities, and detective skills, who resorts to intimidation in fighting crime. Likewise, we claim, slurs invite a certain perspective, as in Camp (2013), one that is negative in the sense of a demeaning stereotype. And so, a use of a slur invites us to delineate our thinking by the tenets of this negative stereotype. And further, even though, as we claim, a speaker avoids objectionable content in using a slur, it is still right to object to, and resist, the objectionable tone. What we object to is not an asserted or presupposed objectionable content but the offensive frame of mind a confrontation with the slur can provoke. We want to resist any contact with this psychological state.

Though some authors agree with us that slurs carry a negative perspective, not everyone agrees with us that this perspective encouraged by a slur is *not* conventionally signaled. For example, according to Camp, 'slurs conventionally signal a speaker's allegiance to a *derogating perspective* on the group identified by the slur's extension-determining core' (2013, 331). But we see no reason to assume perspectival thought is implicated in all tone, or even in all slurs. More importantly, Camp's appeal to conventions frames perspectival thought in *semantic* terms, and we adamantly reject this; we instead acknowledge that audiences readily draw on perspectival thought when prompted by salient experiences, associations, or social practices. There are, moreover, other sources of tone; e.g., think of the metaphors that ground many literal meanings.

Normally, we ignore any imaginative force still lurking in dead metaphors. But sympathetic speakers—and sympathetic listeners—often provoke insights by taking such metaphors seriously. Here is a favorite illustration from Cathleen Schine of what we have in mind:

> I am often accused of 'flying off the handle.' What does that mean? It used to mean, to me, that some member of my family was insensitive,

> unsympathetic, uncooperative and unsupportive. Now, I see myself flying through the air, flung from the handle of an ax like a loose blade, sparkling steel cutting through the blue of the bright sky, soaring, noble and alone, toward the heavens! My life has been considerably enriched. (Schine 1993)

As we can see here but what should be obvious, metaphor involves its own kind of perspective: the use of the imagery of one thing ('the vehicle', ax) to draw attention to analogous attributes of another ('the tenor', speaker). This psychological process is open-ended and non-propositional (Camp 2006), but, as we have argued elsewhere, is not semantic or pragmatic (Lepore and Stone 2015).

The imagery associated with metaphor—especially of so-called stock metaphors—often frustrates a creative and accurate understanding of the vehicle or the tenor; the imagery trades merely in familiar stereotypes. As Max Black (1954) observes, to understand an utterance of 'Sandy is a gorilla' isn't a matter of applying one's factual knowledge (gorillas are highly emotional and intelligent vegetarians living in close-knit family groups). The point is to connect Sandy to a presumption that gorillas are dangerous—what Black calls 'a system of associated commonplaces' irrevocably implicated in the imagery the metaphor invites.

What, then, when a metaphorical tone is demeaning? However else such utterances invite hearers to respond, they more likely than not embody an invidious and ignorant comparison. Even with a dead metaphor, an audience might well be alert and sensitive to these interpretive possibilities—especially if the word applies to them.

Perspective taking and metaphor are only two interpretive strategies that we and others have studied. Camp (2008), e.g., also considers *telling details*, evocative and affectively-laden ways to set the scene and deepen the audience's appreciation of the situation described. As we ourselves (Lepore and Stone 2016) describe *poetic interpretation*, as, in part, an exploration of the evocative potential of the articulation of language itself. The list obviously goes on. Repetition might reduce any of these effects to cliché—'rain' or 'falling leaves' as telling details; 'languorous' or 'brooding' as poetic sounds—or to a potential for tonality available to a sophisticated listener. But what if it is the case that such tone plays into stereotypes and prejudices? Shouldn't we be suspect of words whose loaded associations evoke offensive imagery in such ways?

Furthermore, we mustn't forget that words often evoke the histories of who has used them most notably. These uses can become a kind of

quotation that summons an attitude or milieu the speaker identifies with. One of our favorite examples involves the hipster who adopts the vivid but outdated slang of 1930s and 1940s. He refers to error-prone workmen as 'numbskulls', addresses strangers in casual encounters as 'mister', or announces monetary values as so many 'clams'. The hipster's words, irretrievably tied to an extinct but influential culture, invite an ironic engagement. To catch the hipster's tone is to treat the hard-boiled Los Angeles of film noir or the slapstick New York City of Vaudeville as a model or foil for our current circumstances. As in Schine's metaphor, such attention to the nuance of language promises to assuage our everyday indignities—reframing them as vignettes in a timeless and cathartic drama.

The history of slurs, of course, is not so benign. Much like the Swastika and the burning cross, there are words whose most evocative users waged war on their referents. The link between these words and the ideologies and practices of oppression matters (Hom 2008), though, in our view, it is an historical fact, and surely *not* a semantic or pragmatic one. Just like our hipster, speakers who use these words take antecedent speakers as a model for current circumstances. They threaten to resume old wars—or invite sophisticated listeners to conclude they might. No wonder their words are the most offensive. But, why, the question persists, should we think tone escapes the category of meaning altogether?

Richard (2008) takes the kind of thought and attitude associated with slurs to be part of *what's said* in the utterance of a slur. And Camp (2013), as we mentioned, invokes the perspective-taking tone of slurs with the aim of uncovering *meanings* that distinguish slurs from neutral counterparts, as well as from other slurs for the target group. We disagree with both, and, in some sense, our view is merely an elaboration of Frege and Picardi.

Frege writes, '… coloring and shading are not objective and must be evoked by each hearer or reader according to the hints of the poet or speaker' (Frege 1897, 30).[3] Tone involves guesswork, and '[w]here the main thing is to approach what cannot be grasped in thought by means of guesswork these components have their justification' (Frege 1918, 22–23). When tone is in play, an utterance 'is meant to have an effect on the ideas and feelings of the hearer' (1897, 139). It 'is often said to act on the feelings, the mood of the hearer or to arouse his imagination' (1918, 22). Picardi (2007, 495) speaks of the tone of an expression as casting 'a different light' on its subject matter. In short, there's

something qualitatively different between semantics—the propositional information speakers make public—and tone—the open-ended strategies speakers use to shape one another's thoughts and feelings. This difference is a key theme of Lepore and Stone (2015), where we discuss the poet's provocations for the audience to pursue its own imagery (cf. also, Lepore 2009). Here we offer a brief overview of considerations we think clearly distinguish the two, considerations that we believe substantiate our intuitive assessment of tone as non-semantic (and non-pragmatic) on theoretical grounds.

3 Coordination

Conventional content, together with the collaborative rationality of all joint activity, allows interlocutors to use utterances to contribute propositional information to a conversation. A key feature of linguistic content is that it is *objective* and *publically retrievable*. When we say, 'It's raining,' we normally contribute that it's raining to the conversation, or at least that we think it is. We might also be contributing—more indirectly—that we think our addressee should bring her umbrella. The semantics and pragmatics of discourse work together to explain how these contents become contributed. When we say, 'That was a great lecture,' contingent on the prosody we use to articulate our utterance, an audience may come away thinking we believe the lecture was great or that we are speaking sarcastically and intend to contribute 'the opposite'. There is a matter of some dispute in the literature over whether the latter inference is fixed by the semantics of English or the pragmatics of collaborate conversation. But one way or another, we usually succeed in *retrieving* what our interlocutor is trying to contribute to our joint conversational record.

It's traditional to cash out this idea of the retrieval of any sort of meaning—semantic or pragmatic—in Gricean terms (Grice 1989), that is, meaning must be calculable: it must be capable of being worked out on the basis of (i) the linguistically coded content of the utterance, (ii) the Gricean (1989) Cooperative Principle and its maxims, (iii) the linguistic and non-linguistic context of the utterance, (iv) background knowledge, and (v) the assumption that (i)–(iv) are available to the participants of the exchange and they are all aware of this. However, when we need to tease apart the content of utterances from what speakers merely reveal, prompt, or invite, then Grice's framework becomes quite difficult, if possible at all, to apply. It is better, in our view, to focus on

the distinctive role for coordination in communication, following Lewis (1969).

Coordination can occur when agents face a coordination problem. These sorts of problems crop up wherever there are situations of inter-dependent decision by at least two agents, where coincidence of interest predominates, and where there are at least two coordination equilibria, i.e. at least two ways participating agents can coordinate their actions for their mutual benefit. Agents solve a coordination problem when each acts so as to achieve an equilibrium. They do so by coordination when, if confronted by multiple options for matching their behaviors, they exploit their mutual expectations in settling on one equilibrium (where each agent does as well as he can given the actions of others) to the exclusion of all others.

Lewis illustrates this sort of practice with Hume's example of two men, A and B, in a rowboat: to move, they must coordinate their rowing patterns. There are almost a limitless number of speeds at which each can row, but to row effectively, they need to settle on a single speed, which, interestingly, they can achieve without an explicit agreement. They may stumble on to it; or one might mimic the other. But, should A row at a certain speed because A expects B to do so; and should B row at a certain speed because B expects A to do so; and so on, such that each does his part because he expects the other to do his, then they, thereby, reach an equilibrium through coordination.

The practice of updating the conversational record so as to register specific information also poses a coordination problem. After all, there is no non-arbitrary connection between an utterance and what a speaker can use it to register on the conversational record (other than that the speaker made the utterance). But if the speaker's strategy is to use a particular utterance to get his audience to register particular information, and if he expects his audience to respect this strategy, and if the audience should happen to respect a corresponding strategy in tracking the information that the speaker is attempting to place on the record, and if the audience expects the speaker to respect this strategy, and so on, then the speaker and audience will end up, through coordination, with identical updates of the conversational record.[4]

The way in which agents reliably solve coordination problems is by adhering to a particular scheme implicit in their tendencies or mutual expectations. The key to understanding how coordination functions in solving coordination problems is to appreciate the surprisingly

underappreciated role that conventions play. A convention is a regularity observed by agents, but, of course, not every regularity constitutes a convention; eating and breathing are regularities we each follow but they are not conventional. Someone adheres to a convention just in case his reason for acting in accordance with a certain equilibrium solution to a coordination problem is that he expects others will act in accordance with this same solution to the problem, and that they will do so only if they expect him to act in accordance to the same solution, and he further has some reason for expecting them to act in accordance to the same solution (Lewis 1969, 42). A group of agents are said to share a convention, then, just in case each member does his part in regularity X because she expects everyone else in the group to do their part in X, and each party prefers to do their part in X conditional upon others doing so. Had anyone expected everyone to do his part in another alternative pattern Y, she would have done her part in regularity Y (and not in X).

A convention, in short, is simply a self-perpetuating solution to recurring coordination problem. A group is reliably good at solving a coordination problem only if its members either share patterns of behavior or background knowledge that enables them to choose one pattern over viable others. Since interlocutors are apt at retrieving contributed information from heard utterances, and since each conversation creates a coordination problem for its participants, it follows, by definition, that the participants are exploiting linguistic conventions.

Meaning, understood in terms of coordination, involves a shared effort to demarcate the issues that matter and to adjudicate the answers. Multiple alternative solutions are possible, and so, mutual expectations are decisive in settling the meanings communicating agents agree upon. And in this regard, coordination offers us a clear grounds for distinguishing tone from meaning.

As Frege says, sometimes we 'approach what cannot be grasped … by means of guesswork.' When tone exploits presumed stereotypes or perspectives in a domain, or a creative elaboration of imagery, or the social and historical associations a term has, then listeners draw on their prior experience, and *not* on the expectations they share with the speaker. What's key here is that the results need not be undermined should the speaker not anticipate them in advance. A common theme of our examples of tone—mother and mommy, flying off the handle, falling leaves, brooding, numbskulls (and slurs)—is how a listener's ear for language can develop a feel for an utterance that complements or transcends

the speaker's intentions. Such extensions can amplify on the *point* the speaker is trying to get across and the listener may never fully capture the speaker's point. But, of course, this means that the interlocutors are not coordinating on these unanticipated results.

Reaction to a metaphor, for example, is an open-ended process. Someone calls Bob 'a bulldozer'. What's her point? Suppose an interlocutor amplifies with, 'Oh, yes, he just flattens anyone who disagrees with him'. And the first speaker replies, 'That's a good one'—indicating that she herself had not thought of that particular framing of Bob's actions, but agrees that it captures part of the spirit of her comment. Can it still be part of a correct interpretation? And if so, when would interpretation come to an end? In short, with some contributions, like metaphor, interpretation is genuinely open-ended. We want to belabor this point because we think it's crucial to distinguish open-ended contributions from retrieved ones.

Content by its very nature, must be publicly recoverable; after all, it can be shared. But what could it mean to share open-ended contents? Regardless of how much of the content is—explicitly or implicitly—shared, it would always remain open whether interlocutors share it all. Anyway, such alignment would be unrealistic if interpretive elaborations involve associations derived from individual experience. Meaning is just not the sort of thing that can be left to hints and guesswork. Yet we are not denying that open-ended interpretive effects are important—they do 'have their justification,' as Frege writes. They can be prompted by the speaker's choice of specific words. A sensitive speaker can invite a sympathetic reader to get more precise insights through any of these associations. Conversely, a hearer may run with any of these interpretive strategies, regardless of the speaker's intent, simply because they strike her as apt. This is tone, which, by its very nature, can work in problematic ways—reminding us of difficult histories, or revealing the speaker's prejudices, or evoking upsetting situations as models for our present interactions, inviting demeaning perspectives. Such tones can be objectionable; we may even prohibit the corresponding words as slurs.

4 Conclusion

We have argued that interpreting a slur is not the same as ascribing meaning to it. Instead, the reason why slurs can infect our judgments is because of the particular way of thinking that they provoke. And one

reason for this is because slurs are prohibited: their utterances are deliberate violations that we cannot help but react to. But slurs, we argued, can carry a wide range of tone, and so, provoke a wide range of effects. Nevertheless, to have any hope of distinguishing determinate propositional contributions from the guesswork of open-ended, improvised rapport, we *have to* exclude tone from meaning. In short, slurs have no special meanings.

Granted, people convey meaning with language, but they use language in other ways for other purposes. Elsewhere we suggested principled ways of distinguishing the meaningful contributions associated with utterances from their imaginative insights (Lepore and Stone 2015). And we offered principled ways of understanding how interlocutors pursue these contributions and insights collaboratively, both to reach agreement on how things are, and to share their thinking and perspectives. The approach involves more distinctions about meaning and agency than philosophers have typically made, but it's crucial to draw them. They are needed to reconcile ordinary intuitions about language with the rich interpretive landscape that we live in.

We emphasize that none of what we say here denies slurring has a *point*. A slurring utterance is not merely an affective or expressive act that displays negative feelings for its target. However, even though an individual may intend to get some point or other across with a slur, what the audience actually takes away from it is complex and indefinite. When taking in the point of a slur, we draw upon a particular kind of engagement with the speaker's utterance and the world she describes. We are using the speaker's language as a prompt. Such aims do not seem propositional in nature. There does not seem to be anything specific the slurring utterance means. Any utterance can make contributions to discourse, reveal its speaker's assumptions and suggest a particular perspective on its subject matter. Recognizing that an utterance is exhibited for a certain kind of effect gives the audience's understanding a new interpretive twist.

In some cases, we can explain the prohibition of a slur in terms of such interpretive twists. In particular, the explanations of taboos against slurs found in Hom (2008) or Camp (2013) remain possible on Prohibitionism. Of course, there are many reasons for terms to be prohibited (Anderson and Lepore 2013a). We need not always justify prohibitions via differences in meaning. But if a term is loaded with certain kinds of tone, a speaker who uses it, wittingly or not, foists on an audience something that can be quite difficult to reject, something resistant

to direct challenge. Engagement with a slur potentially renders an audience complicit in the response it provokes, even when they find it repugnant. The prohibition against its use is as attempt to protect an audience from the workings of their own psychology, from the potential infections slurring can cause.

Acknowledgements An earlier and extended version of this paper was presented at the Philosophy Department, Peking University. Thanks to Elisabeth Camp, Bob Carpenter, Sam Carter, Doug DeCarlo and Mitzi Morris for comments and discussion on its topics. Preparation of this research was supported in part by NSF grant IIS 1526723 and a sabbatical leave from Rutgers to Stone.

Notes

1. Cf. Frege (1918, 22–23) for differences in tone between 'horse' and 'steed'.
2. So, we depart from Dummett (1981), for whom tone is part of language, and so, objective and public; though, as he notes, '…there is no reason to suppose that all those variations in meaning between expressions having the same sense…, which Frege counts as difference in tone, are uniform in kind' (Dummett 1981, 85; also, 1991, 122; Picardi 2007, 500). We believe Dummett and Picardi agree with us that tone may invoke psychological associations, but have been misdirected because of all the phenomena that Frege included under tone, e.g., differences between 'and' and 'but'—which are clearly conventional.
3. 'Farbung,' and also, 'Beleuchtung' (lighting or shading), and 'Duft' (fragrance or scent).
4. Of course, there needs to be mutual recognition as well. See Lewis (1969).

Bibliography

Anderson, Luvell, and Ernie Lepore. 2013a. What Did You Call Me? Slurs as Prohibited Words. *Analytic Philosophy* 54 (3): 350–363.

Anderson, Luvell, and Ernie Lepore. 2013b. Slurring Words. *Noûs* 47 (1): 25–48.

Black, Max. 1954. Metaphor. *Proceedings of the Aristotelian Society* 55: 273–294.

Camp, Elisabeth. 2006. Metaphor and That Certain 'Je Ne Sais Quoi'. *Philosophical Studies* 129: 1–25.

Camp, Elisabeth. 2008. Showing, Telling, and Seeing: Metaphor and 'Poetic' Language. In *The Baltic International Yearbook of Cognition, Logic, and Communication, vol. 3: A Figure of Speech: Metaphor*, 1–24. Manhattan, KS: New Prairie Press.

Camp, Elisabeth. 2009. Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments. In *Midwest Studies in Philosophy: Poetry and Philosophy*, 107–130. Oxford: Wiley.

Camp, Elisabeth. 2013. Slurring Perspectives. *Analytic Philosophy* 54 (3): 330–349.

Dummett, M. 1981. *Frege: Philosophy of Language*, 2nd ed. London: Duckworth.

Dummett, M. 1991. *The Logical Basis of Metaphysics*. London: Duckworth.

Frege, Gottlob. 1892 (1952). On Sense and Reference. In *Translations from the Philosophical Writings of G. Frege*, ed. P. Geach and M. Black, 56–78. Oxford: Blackwell.

Frege, Gottlob. 1897 (1979). Logic. In *Posthumous Writings: Gottlob Frege*, ed. H. Hermes, F. Kambartel, and F. Kaulbach, trans. P. Long and R. White. Oxford: Blackwell.

Frege, Gottlob. 1918/1919. The Thought: A Logical Inquiry. In *Basic Topics in the Philosophy of Language*, ed. R. Harnish, 517–535. Englewood Cliffs, NJ: Prentice-Hall.

Frege, Gottlob. 1979. *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach, trans. P. Long and R. White. Oxford: Basil Blackwell.

Grice, H.P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hom, Christopher. 2008. The Semantics of Racial Epithets. *Journal of Philosophy*. 105 (8): 416–440.

Horn, J. 2013. Mommy Wars. *Kveller*.

Lepore, E. 2009. The Heresy of Paraphrase: When the Medium Really Is the Message. *Midwest Studies in Philosophy* 33: 177–197.

Lepore, E., and Matthew Stone. 2015. *Imagination and Convention*. Oxford: Oxford University Press.

Lepore, E., and Matthew Stone. 2016. The Poetic Imagination. In *Routledge Companion to Philosophy of Literature*, ed. J. Gibson and N. Carroll, London: Routledge.

Lepore, E., and Matthew Stone. 2017. Pejorative Tone. In *Bad Words*, ed. David Sosa. Oxford: Oxford University Press.

Lewis, D. 1969. *Convention*. Cambridge, MA: Harvard.

O'Hehir, Andrew. 2015. So Much for Youth Apathy: Student Radicalism Escapes the 60s at Last. *Salon.com*, Nov 17. Online at http://www.salon.com/2015/11/17/so_much_for_youth_apathy_student_radicalism_escapes_the_60s_at_last/.

Picardi, Eva. 2007. *'On Sense, Tone and Accompanying Thoughts', the Philosophy of Michael Dummett*, 491–520. Chicago: Open Court.

Richard, M. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.

Schine, Cathleen. 1993. ON LANGUAGE; Dying Metaphors Take Flight. *New York Times*, August 8.

Refusing to Endorse: A Must Explanation for Pejoratives

Carlo Penco

1 Introduction

Since David Kaplan's "The Meaning of 'Ouch' and 'Oops'", there has been a wide amount of discussions on every side of pejorative expressions or slurs, with different kinds of interpretations and new topics, like the problem of appropriation and perspectival shift.[1] Picardi (2006, 2007) presents a set of suggestions concerning the use of pejoratives and their relation to the content of what is said. Her stance is antagonist towards a too easy "pragmatic" view of the matter, according to which a pejorative doesn't touch or is totally independent of what is said and only pertains to the level of implicatures or presuppositions. On the contrary, Eva claims that the use of a pejorative cannot be reduced to something always independent of the assertive content, and that the use of pejoratives may pertain to the truth of the matter, given that it predicates something false of the class to which it refers. Therefore, she would be classified as belonging to the "semantic stance" proposed for instance by Hom (2008, 2010, 2012). According to the semantic stance, the

C. Penco (✉)
University of Genoa, Genoa, Italy
e-mail: penco@unige.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_10

derogatory content is part of the meaning of the pejorative (e.g. "nigger" *means* something like "black and despicable because of it"), and therefore a sentence containing a slur attributes an empty property to the individual in question (Picardi 2006: 72), making the sentence either false or deprived of truth value. Although she claims that the derogative aspect of pejoratives is "part of a word's literal meaning", I think the morale of her papers points towards a wider view on the role of pejoratives than the semantic one. I will follow Eva's analysis of multi-proposition view (§1), her attempt to make derogative terms impinge on truth conditions (§2), her reaction of Tim Williamson's criticism of Michael Dummett (§3), her dubious attitude towards a presuppositional analysis (§4) and eventually, in (§5), I conclude with a solution that seems to prompt from her discontent with most answers to the problem of derogatory terms.

2 Sense, Tone and Accompanying Thoughts: A Multiple Propositions Analysis

In order to distinguish what a pejorative expression add to what is said, Kaplan (1999) distinguishes *descriptives* and *expressives*: the first *describe* what is or is not the case; the second *display* what is or is not the case under a certain perspective or attitude (two expressions may have the same information content and different expressive content). According to Kaplan himself, this distinction is not so distant from the Fregean analysis on the different contribution to content made by sense and tone. Picardi (2006, 2007) discusses the Fregean distinction in relation to the use of derogatory words. Frege considered tone or colouring as of pragmatic significance and not pertaining to the truth-conditional content of an assertion (the assertoric content). Frege's distinction antedates the distinction made by Paul Grice (1967) between what is said (truth-conditional meaning) and what is meant (conventional or conversational implicatures), as Picardi (2001) was one of the first to remark. A standard example is given in Frege (1897) analysing the difference between:

1. "That dog howled the whole night"
2. "That cur howled the whole night"

According to Frege, the two sentences express the same truth-conditional content: if the first is true, then the second also is true.

Frege claims that (2), although expressing the disapproval of the speaker, cannot be false if (1) is true. In fact, if we thought that the aversion of the speaker was *part* of the content, the sentence should be analysed as a conjunction of (1) "that dog howled the whole night" and something like (3) "all dogs are despicable and ugly". Assuming that (3) is false, the conjunction of (1) and (3) would be false. Therefore, we could not accept the truth of sentence (2) given that it is an expression of a complex proposition whose truth-value is false. But we cannot assume that (2) is false while (1) is true, given that they refer to the same state of affair. Supported by this argument, Frege distinguished between thoughts whose assertion is expressed and thoughts that are not expressed, but only hinted at or "suggested", in order to influence the audience. Suggesting something using a particular piece of the lexicon to refer to an individual does not concern the problem of truth and belongs to the realm of colouring or tone, which pertains to pragmatic aspects of language (Frege 1879: §3, 1897). The solution to the analysis of pejoratives seems straightforward: conveying something suggested and not explicitly asserted, a sentence with a pejorative does not concern what is said, but what is meant, or the conventional implicature. Picardi is not happy with this solution and tries different ways to go beyond it.

First of all, Picardi criticises Frege for assuming too easily that "dog" and "cur" are coreferential. If so, the two terms should require substitutability *salva veritate*, but there are counter examples:

> To his neighbor's utterance 'That cur howled all night', the owner of the dog may retorts, 'That dog is not a cur', but plainly he is not asserting that his dog is not a dog. Possibly, all curs are dogs, but not all dogs are curs. All that Frege is entitled to say is that there are contexts of utterance in which the difference in meaning between "cur" and "dog" makes no difference to truth-conditions of what is said, whereas there are other contexts in which the difference is salient. (Picardi 2006: 62)

The main claim given by this example is that we cannot take it for granted that a neutral term and a pejorative have always the same extension. However the disagreement between two speakers here does not grant the conclusion; in fact the two expressions (the neutral one and the derogative one) have the same extension in the mind of the dog hater, and when they are used to refer, the reference is normally successful because the interlocutor easily gets what the speaker has in mind.

Saying that it is false that all dogs are curs is an expression of disagreement on the different uses of the words, not on the factual truth of the assertion of the speaker (whether the animal—in whichever way you want to refer to it—howled all night). It seems to me therefore that this argument is not strong enough to avoid the conclusion that assertions containing pejoratives have the same truth-conditional content than assertions with neutral terms, insofar the pejoratives are used to refer and are understood as such.

What about a multi propositional analysis? Relying on a long tradition of research (Kent Bach, Robyn Carston, Francois Recanati), Picardi claims that the difference between "what is said" and "what is conventionally implicated" is not sharp enough to decide without doubts when something belongs to the content of an assertion and when it does not, given that the choice may depend on the *point* of the assertion. A possible solution might be to translate the sentence (2) with an explicature (or a free enrichment), as:

4. "That dog, which is despicable and ugly because of it, howled all night".

With this peculiar rendering, we might answer to the Fregean strategy for which it is counterintuitive to take (2) as false considered as a conjunction ("that dog howled *and* all dogs are despicable and ugly"). In fact, with (4) interpreted as an explicature of (2) we would really have a different proposition from (1) and we may admit—in this case—that the truth of the content of the main assertion (that the dog howled all night) is affected by the truth-value of the relative clause. This might be a possible "multi-propositional" solution of the relevance of pejoratives to what is said.

We may claim that the use of the pejorative is intended to imply that the fact that dogs are despicable is a *reason* or a cause why they howl all night or vice versa. This last point seems the best way to explain Picardi's criticism of Kaplan's analysis of Frege's accompanying thoughts (*Nebengedanke*) with which she shares much, but not all. The discussion starts with Frege's example:

5. "Napoleon, who recognized the danger to his right flank, personally led his guards against the enemy position"

Picardi remarks that Frege realizes that "the clause expresses more through its connection with another than it does in isolation" (Frege 1892: 47). A relative clause cannot be represented always with the same syntactic form. We have different ways to compose an accompanying thought expressed by a relative clause, and while some of them make it independent of the whole, others really affect what is said with the composed sentence. In the case under discussion, Picardi suggests that what is said may be conceived as inserting a third thought, that is

6. "the recognition of the danger is a reason why Napoleon led his attack".

In this case, it becomes apparent that the accompanying thought may be part of what is said, given that it impinges on the truth-condition of the composed sentence. Picardi really makes a case about that. The conclusion is that a pejorative may be treated as prompting a further proposition that cannot be conceived just as a conventional implicature, but as an explicature—that is part of what is said—presenting a point, such as individuating the reason explaining the content of the main clause: seeing the danger of his side is a reason for Napoleon to attack, or being ugly and despicable is a cause for the dog to howl all night—maybe because despicable animals just do that.

Although this is a possible analysis of pejoratives that makes justice of the idea that an assertion containing a pejorative may be just false, Picardi eventually rejects it. In fact she claims that the idea of a *specific* completion of a sentence because of a pejorative is not sound; following Sainsbury (2001), she claims that what is relevant with a sentence with expressive content is its *lacking of specificity*, and therefore the sentence "should not be construed as directed to an elliptical proposition that awaits to be spelled out in full" (Picardi 2006: 54).

3 Do Pejorative Really Impinge on Truth Conditions?

Without the help of explicatures or free enrichment, however, it becomes difficult to claim that pejoratives pertain to the assertoric content, to what is said. Yet Eva, criticising Kaplan's too sharp separation between expressives and descriptives, attacks the rendering of this distinction made by Potts (2005), who considers expressives as conventional implicatures. According to Potts the expressive meaning of a lexical item

is *independent* of its descriptive meaning and therefore plays no role in determining the truth conditions. The main point of disagreement with the above distinction concerns the claim of *independence* of the expressive content. Eva's criticism of the idea of independence of expressives works on a basic question:

> How can we consider the *correctness* of a reported speech in case the original speech contained a pejorative?

At first sight, reporting an utterance with a pejorative like "that cur howled all night", a lover of dogs would probably make a report of the kind: "*x* said that that dog howled all night", abstaining to use the pejorative term, but still thinking to have made a correct report of what happened, preserving at least the truth of the matter. But not everybody would agree of the correctness of the report. Eva refers to Bach (1999) whose argument for claiming that conventional implicatures belong to what is said is that they fail the indirect speech test; if you report John's having said "Mary is pretty *but* intelligent" as "John said that Mary is pretty *and* intelligent", Bach doubts that you have made a correct report. In the reported speech we should make it clear that the speaker intended a contraposition between the first and the second property. We should have an enriched proposition that could make explicit the content of the contrast. We have seen however that this is not the path followed by Eva. Which means are still available to fight a analysis of pejoratives based on the idea of conventional implicature?

Instead of following the multi-propositional analysis, Eva pinpoints another possible problem: the relevance of what is the "at-issue" content or the question under discussion (QUD). On this point she makes an example purporting to show the difficulty of sharply separating the assertoric content from the implicated content:

> Whether my leaving out this piece of information renders my report wrong or simply inaccurate depends on what was the main point of the utterance on the given occasion. And this, in its turn, depends on the audience I am addressing and on the focus of the conversation: in the course of an investigation that aims at discovering the culprit of evil deeds against dogs in a certain neighbourhood it may be useful to give a literal report of what the people involved say concerning dogs. In a different context, the report may be less accurate, if, for instance, our interlocutor simply wants to find out what a notoriously nagging neighbour was complaining about.

This is an ingenious effort to defend the claim that pejoratives enter the question of truth. But it seems to me that here we have two questions: one question concerns the facts described by the report, another question concerns the facts concerning the psychology of the speaker: if we are looking for a devious assassin of dogs, reporting the specific lexical item impinges on the latter. We might have evidence, although inconclusive, of the speaker's tendency to perform crimes against dogs. Here, therefore, a literal report may be of fundamental importance to denounce the speaker. However, again, derogatory conceptualizations do not change the "strict" truth-conditional content of a description of a state of affairs. In fact, in this case, the truth evaluation concerns (the fact of) which *words* the speaker used, not which *facts* have been reported about the behavior of dogs.

In fact, "that cur" is a complex demonstrative, whose *main* role in the sentence is the identification of the referent; we may think that it presents the referent in a wrong way and, from this perspective, is not too dissimilar to a misdescription. A speaker may make a referential use of an inaccurate definite description assuming the hearer may understand the intended referent although the description is false of it, or at least defective (see also Penco 2017a). From the point of view of truth conditions, both misdescriptions and pejoratives may be considered defective but still able to make the hearer correctly understand the referential intentions.

Picardi (2006: 67) is well aware of the problem, and she refers to Donnellan on this point. Her use of the similarity with Donnellan's cases helps to point out the *difference* between the case of misdescriptions and the case of pejoratives. In case of misdescriptions like "the man drinking champagne" (while he is drinking mineral water), there is no harm in using a defective or inaccurate or wrong definite description if your referential intentions are understood. On the contrary this does not happen with pejoratives. While with misdescriptions we are in front of a factual mistake, whose correction is easy to accept ("the person you are referring is not drinking chanpagne but mineral water"), in front of a derogatory term you cannot simply change the term and be happy, because you are facing a strong disagreement: what is wrong from the point of view of dog lovers, may be strongly believed by the dog hater, who would not recede from his conceptualization of that class (dogs are despicable because of being dogs and blacks are inferior because of being black). Besides, given certain circumstances, I may easily report what a speaker

said using the same misdescription to make myself understood, but this is not so with pejoratives. In fact I might feel uncomfortable using a pejorative term, on whose grounds and consequences I don't agree. But this does not mean that I don't *understand* what the speaker said and I have to distinguish between the facts of the matter described and the point of view and attitude of the speaker.

As Dummett (2007: 527) says, commenting of Picardi's paper, "the use of a pejorative expression certainly cannot be said to affect the truth-value of an utterance; it affects its property. But, for the same reason, it also cannot simply be explained as affecting the tone of the utterance, or as attaching an implicature to it." The offensive character of certain terms, Dummett claims, should be accounted for by "the license they give their user to draw inappropriate consequences". In conclusion, we cannot use the test of reported speech to claim that pejorative impinge on the truth of the matter, although we may still take our awareness on "the tacit commitments we would undertake in accepting a certain way of referring to people or actions" as a ground to refuse to endorse an assertion (Picardi 2007: 507).

The question seems to shift from the truth-value of an assertion concerning a state of affair (what did the dog do during the night?) to the justifications and consequences of assertions containing a pejorative. If truth conditions are not affected, pejoratives certainly affect assertibility conditions. Different lexical items are connected with different justifications and consequences, and using them obliges us to explain why we have used them and commits us to the consequences of what they mean. We are entering another kind of problem, that touches upon the meaning of pejoratives: while it seems that truth conditions are affected only by the *referents* of pejoratives, assertibility conditions may be affected by their *meanings*. On the meaning of pejoratives, Picardi is very near to Dummett's classical analysis and contrasts Williamson's criticism of this analysis.

4 Pejoratives as Dealing with Truth/Assertibility Conditions

Picardi (2006, 2007) looks back at Dummett's discussion on the logic of pejoratives. Dummett was interested in the logical role of pejoratives and in the logical motivations to reject their use. His claim was clear and

simple: a pejorative like "Boche"—used to refer to Germans implying that Germans are more prone to cruelty than other Europeans—would produce a non conservative extension of the language where the word was not present: the use of "Boche" would permit inferences and conclusions that would not be permitted in the language missing the pejorative (Dummett 1981: 454).

Dummett gives Introduction rules and Elimination rules for the term and shows how they permit conclusions impossible to be derived without the term. The Introduction Rule for "Boche" (or its condition of application) would simply be something like

x is German

———————

x is Boche

But the consequences of application embed the following inference (that might be considered the Elimination rule for "Boche"):

x is Boche

———————————————

x is more prone to cruelty than other Europeans

Now, if we accept the Introduction and Elimination rules for *Boche* we should derive something of the following: Angela Merkel is German therefore is *Boche*, and if Angela Merkel is *Boche*, therefore, she is more prone to cruelty than other Europeans. We could not derive this conclusion from our lexicon only if, following the elegant attitude of Oscar Wilde, we did not include the lexical item "Boche" in our dictionary.

Williamson launches an attack on radical inferentialism and defends a radical referentialist framework (although he recognizes that there are intermediary positions that might escape his criticism). He criticizes Boghossian (2003: 241–42) according to whom "plausibly, a thinker possesses the concept *Boche* just in case he is willing to infer according to [Dummett's rules]", with the following short argument:

> Since understanding the word 'Boche' (with its present meaning) is presumably sufficient (although not necessary) for having the concept that 'Boche' expresses, it follows that a willingness or disposition to reason according to Dummett's rules is equally unnecessary for having that concept. (Williamson 2009: 8–9)

This claim is correct, and maybe Boghossian went a bit too far. Let us assume that understanding a concept is understanding its introduction and elimination rules. Mastering those rules is not to be identified with willingness to follow them, but with an implicit knowledge of them. Understanding the meaning is understanding what it is or *what it would be using* that inference, even without explicitly doing so or even rejecting to endorse it. Williamson may accept that, but then—he would ask—which is the difference between an inferentialist and a referentialist account of understanding as a "practical" ability if we cut off the actual disposition to reason according to the rules? The answer is that, although there is no difference in "practical" ability, inferentialists are not content of getting the referent right: they require making the inferential connections explicit. Explicitly rejecting to use a term is exactly the point of the difference between a referentialist and an inferentialist view. You refuse to use a term because you reject the possible consequences of its use; referential rules on the other hand just point out that the reference of "German" is the same of the reference of "Boche"; the two classes are the same class:

> differences between 'Boche' and 'German' apparently play no role in determining reference, and so make no difference to the way in which the terms contribute to the truth-conditions of sentences in which they occur, a Fregean might even count 'Boche' and 'German' as having the same sense. Frege himself gives just such an account of another pejorative term: 'cur' has the same sense and reference as 'dog' but a different tone. (122)

Here we are. Again on Frege, and our examples of pejoratives for "dogs"! As we have seen, the main role of pejoratives does not concern just the role of reference fixing, as in case of definite descriptions, but their role in suggesting inferences to be accepted (conventional implicatures). On this point, Williamson himself concedes the idea that conventional implicatures (something derivable and therefore linked to an inferential structure) of expressions like "Boche" are "part of their meaning in a broad sense of meaning". But, if we accept an idea of (broad) meaning as dealing with the inferences connected with an expression, then Dummett's proposal is not incompatible with a view of pejorative as triggering a conventional implicature or a presupposition.

A way to interpret Dummett's treatment of pejoratives is then to consider it as a clarification of the rules behind what is expressed and not stated, rules that should be followed if one accepts the conventional implicature connected with the use of the pejorative. Accepting a pejorative, we accept a network of inferences, a set of beliefs that the pejorative brings with it. Using an assertion with a pejorative is not only saying something true with a bad psychological surrounding: it is accepting the consequences connected to the inferential meaning of the expression. We are back to the conclusion of the previous discussion: we *understand* the intended referents of singular terms or complex demonstratives like "that dog" or "that cur", and we understand to which classes predicates like "German" or "Boche" refer; although sentences containing them may have, by substitution of coreferentials, the same truth conditions, they certainly haven't the same assertibility conditions; in fact, to have the same assertibility conditions they should also have the same ground for justification, and we may claim, from our perspective, that nobody is justified to use "cur" or "Boche" given that those terms imply consequences that we disagree about.

5 Problems of Pejoratives as Presupposition Triggers

If a conventional implicature can be considered part of the "broad" meaning of an expression, then it seems that Picardi (2007: 508) herself makes a too strong contrast between "the decision to construe the explicit derogatory ingredient as a conventional implicature" on one hand and the idea of construing the derogatory ingredient "as constitutive of word meaning" on the other. The two aspects are not antagonist: considering the derogatory ingredient as conventional implicature implies that its broad meaning is connected with the inferences that are derivable by its use and are suggested as "calculable" implicatures.

However, speaking of inferences syntactically plugged into the lexicon, conventional implicatures may not be the best solution for treating pejoratives. The other solution is treating them as triggering presuppositions. Actually conventional implicatures pass the S-Family test of presuppositions: they stand also when an assertion is made in negative, interrogative and modal form ("that cur didn't howl all night", "did that cur howl all night?", "that cur might have howled all night").

Let us then see what happens when treating pejoratives as presupposition triggers (for a defence of a presuppositional account see Schlenker

2007; Macià 2014; Cepollaro 2015). Under this perspective, the use of a pejorative presupposes the set of beliefs that the pejorative intends to implicate. This should be coherent with the classical view of presuppositions for which an utterance of a sentence is *appropriate* if the presupposition is shared by the participants of a conversation (Stalnaker 2002). We may consider the use of pejorative not appropriate if we do not share the presupposition; therefore, as Picardi (2007: 507) claims "we may abstain from accepting a statement made by others because we are aware of the tacit commitments we would undertake in accepting a certain way of referring to people or actions".

This claim is perfectly adequate to a presuppositional analysis, and it seems to me that presuppositional analysis and the difference between truth conditions and assertibility conditions come hand in hand; abstaining from endorsing a statement means rejecting the justifications or the intended background for its assertion.

A presentation of a presuppositional analysis might also be framed in Kaplan's terminology. Kaplan (1999) was interested in the informational content that can be derived by expressives; in doing so he attempted to clarify the rules of correctness of expressives and the correspondence of informational content given by expressives and by descriptions: the same semantic information can be given with an expressive mode (ouch, oops, hurray) or with a descriptive mode ("I am in pain" or "I have just observed a minor mishap" or "I am in state of joyful elation"). Kaplan describes the problem of giving the rules for correct application of expressives. And we might say that the felicity condition of the use of a pejorative is that (1) the person actually believes the information content expressed by the pejorative and (2) has the correct attitude or emotion towards the class described by the pejorative. Utterances of "that Boche run away" or "that cur howled all night" are appropriate only if the speaker really believes that Germans are cruel as such or that dogs are despicable and ugly as such and has the appropriate emotion of distaste or dislike (see also García-Carpintero 2017). This is what presuppositional analysis amounts to.

However, the presuppositional analysis of pejoratives leaves unanswered some questions like the following:

1. In using a slur in a re-appropriation case, people do not share the prejudice (the belief) attached to the term, yet it seems that their use is appropriate.

2. When people who have racist prejudices use the derogatory term we should say that their use is perfectly "appropriate" because they share the beliefs supporting that use, yet it sounds awkward, although correct for the theory, to say that the uses of derogatory terms are somehow "appropriate".
3. when a presupposition is expressed it ceases to be a presupposition and it is normally accepted while the presupposed content of a slur is typically a matter of disagreement when explicitly stated.

Leaving (1) and (2) to the reader, let us see the problem with (3). Accommodation (the process through which people accept presuppositions that do not belong to the common ground) is not as normal as it is in standard cases (where, as Lewis says, presuppositions "spring into existence making what you say acceptable after all"). A non-xenophobe, or a non-racist, or a friend of dogs, would not easily accommodate the presupposition in a sentence that uses a pejorative. He would probably say, "Hey, wait a moment! Do you think that Germans are more prone to cruelty than other Europeans? It is not true" or "hey wait a moment! Do you think that all dogs are despicable? That's false". The problem arises because the presupposition triggered by a pejorative represents a content on which there may be very strong disagreement.

The main defect of presuppositional analysis is that it leaves something out; offensive or derogatory terms does not only pertain to the *content* of their presuppositions (and eventually the emotional attitude going with it); they also involve actions and commitments undertaken in their use.

6 Pejoratives as Higher Order Utterance Modifiers

Eva makes a remark on Frege's view of the derogatory ingredient attached to the word "cur":

> [according to Frege] in the given context the choice of "cur" instead of "dog" has the value of an exclamation, and, one may add, could be rendered syntactically by means of an exclamation mark, much as assertoric force is rendered by means of a vertical stroke. Frege held that assertoric force only shows itself with the help of a suitable notation, but is not located in any part of speech in particular. Its scope is the whole utterance, not a particular segment of it. The function of an interjection mark

> encapsulated, as it were, into the meaning of "cur" in the specific utterance is to disclose the attitude of the speaker towards the matter at issue. It presents the dog as ugly or unpleasant from the speaker's perspective; however, as Frege remarks, the dog itself may very well be a handsome representative of its race. But this circumstance does not render the use of the interjection incorrect, for in uttering as he did, the speaker might have wished to disclose his attitude of dislike or fear of dogs in general, not of this dog in particular. (Picardi 2006: 62–63)

Eva here refers also to Kaplan, who distinguishes between "truth *simpliciter*" and "truth with an attitude", but—as we have seen—she does not agree to treat the expressive content merely as something propositional: "tone need not be expressible by means of *a*, let alone *one* specific, full-fledged proposition" (Picardi 2007: 503). But which kind of non-propositional aspect can be conveyed by a derogatory expression?

Eva attributes the main reason for accepting the Fregean suggestion of colouring as higher order utterance modifier to the fact that it detaches the notion of colouring from mere psychological significance.[2] A pejorative may impinge on the level of speech acts, on their felicity conditions or justification (or assertibility) conditions. This is a central point to be clarified.

Eva oscillates between two alternatives often connected: a multipropositional account and a higher order account, both of which she tends to disregard. But I think she has been too quick with disregarding the idea of higher-order account, maybe because too strictly connected with the Gricean view. Speaking of higher order account we typically tend to consider the contribution of some words (expressions like "but" or pejoratives) as *parasitic* on a ground floor or central speech act (Grice 1989: 361–62).[3] But the idea of higher order modifier is not exhausted by Grice's view of implicatures (nor by the presuppositional account). An alternative view may be defined for treating pejoratives as higher order modifiers that are not just parasitic aspects. Still keeping pejoratives as connected with a set of inferences (presupposed or implicated), García-Carpintero (2015, 2017) tries to give them a further role in the context of dialogue. The novel point that García-Carpintero makes it *where* to insert the role of pejorative in the dialogue: not only as part of the content or as presupposed propositions, but as constraints on the context of dialogue. The main consequence of accepting derogatory expressions is the implicit acceptance of their presuppositional content, given by tacit

accommodation. Tacit accommodation implies tacit undertaking of a network of inferences and commitment to the consequences.

The main point made by García-Carpintero is, therefore, that common ground cannot be defined only in terms of shared propositions, but also in terms of different commitments towards those propositions; that's why we feel so uncomfortable when we are included in a conversation where people use pejoratives on whose stereotypical inferences we strongly disagree. Already Stalnaker claimed that we have different attitudes towards the contents of the common ground (see Domaneschi et al. 2014). But García-Carpintero point is stronger and can be summarised by the claim that our common ground is made not only of propositions and *propositional* attitudes but it also concerns attitudes linked to *illocutionary* forces, which is a further level of pragmatic commitment.

Saying that the use of pejorative is linked to illocutionary force is a fundamental step, shared by many others. Langton (2012) calls "speech acts of subordination" those speech acts used to classify a target class as inferior, legitimate discrimination and deprive it of rights. But her examples are basically explicit acts of subordination like "Blacks are not permitted to vote", where the act is a kind of "verdictive" and the speaker has authority to do that because he is in a social position that allows him to perform the act. Besides, in case the speaker has no authority, the accommodation of the presupposition (given by the failure to question the speaker) would confer authority to the speaker herself, as suggested by Maitra (2012) (McGowan 2004, 2009 speaks of "conversational" exercitives that, differently from Austin's, do not require uptake from the hearers).

However most of the examples of this literature concern *explicit* and *direct* acts of subordination (like the above quoted "Blacks are not permitted to vote"), or hate speech that is characterised, among other things, by being *directly* addressed to the individuals whom they insult (see also Hornsby 2003: 297). On the contrary our examples (following Frege's example with "cur") concern the use of derogative words in descriptions of facts or in questions, where the pejorative is part of a descriptive content of a phrase (complex demonstratives, definite descriptions) whose main function is to pick a referent. How to describe the subtler way in which the insertion of a pejorative in a normal description of facts changes—to use Lewis' terminology—the conversational score?

A first solution is to think of an *indirect speech act*: by putting a *question* such as "do you know what time is it?" I make a *request*; by

describing a possible situation such as "I will not miss your date" I am making a *promise*. Following the analogy, we may say that by *describing* a situation with "that Boche run away", or asking "have you seen that Boche running away?" I hereby *promote* discrimination and *legitimize* behaviour of discrimination. But think of when the preparatory condition is not satisfied. If asked: "can you tell me the time?" I may answer: "Sorry, I have no watch", making it clear that the preparatory condition of the request is not satisfied. Analogously, if asked: "have you seen a nigger running away?" I may answer: "Sorry, for me there are no nigger", because the preparatory condition of the act of discrimination requires that black people are inferior as such. But the analogy is not so clear: an indirect speech act is typically a speech act of a kind that is given by a speech act of a different kind. By a question we make a request, by a description we make a promise; in case of assertions or questions containing a derogatory term we are still making an assertion or a question. Saying: "hey, wait a moment; he run away, but he is not a nigger" (as with rejecting a presupposition), we correctly answer the main speech act; while we cannot say to a question like "do you know the time" something like, "yes, we do, but unfortunately I have no watch". The strategy of indirect speech act after all seems not to be a viable analysis.

A second possible answer, that seems to be more coherent with the main trend in contemporary discussion, is that speech acts with derogatory terms contain a peculiar *adjunctive force*: with the same utterance, we make two kinds of speech acts at the same time (Kissine 2013: 197): assertions, questions, commands, and other speech acts can be understood as such, and at the same time, when containing a derogatory term, they are *at the same time* acts of "subordination". And also, we have two contents: the (description of the) objective state of affairs (a person who runs away) and the (promotion of a) derogatory viewpoint concerning the individual and the group they belong to. Langton (2017) presented a similar idea at the ECAP Conference in Münich, speaking of "Blocking as Counter-speech" (e.g. you may assert something normal and at the same time, through a presuppositional trigger, convey something else like in "Even John could win!").[4] We may conclude that speech acts that contain derogatory terms (or other subtle means to give a diminishing perspective on the target) promote and legitimize subordination or other negative attitudes towards the referent class. And, most of the time,[5] the subordination is derived by the use of a predicate that is false of the class

in question, for instance, because "the complex properties indicated by racist words are not instantiated" (Picardi 2006: 68).

Can we be content in saying that with pejoratives we make two speech acts at the same time? The idea of a speech act of subordination is still a pointer towards an idea to be refined, and we might distinguish levels of subordination, and also other kinds of acts depending on different kinds of pejoratives or on different targets or different social roles (on which see e.g. Popa-Wyatt and Wyatt 2017). The essential feature, besides the actual contempt or disregard of the target, is normally taken to be that in using a pejorative we act to promote and legitimize subordination. What is not yet clear is *how* this promotion is realised.

My suggestion is that who intentionally[6] uses derogatory terms looks for company, for sharing the prejudice and avoiding feeling alone. His speech act constitutes a *call for joint responsibility*, asking for sharing an attitude towards the derogatory content, indirectly creating a context of commitment to certain behaviour against the target. This is why rejecting to endorse an assertion or answering a question containing a pejorative is the fundamental reaction and avoids the trick of the use of derogatory terms; on the one hand it seems that the racist (or the dog hater) is just stating some facts and therefore we are ready to accept or reject the truth of the matter; but in stating some facts with a certain terminology the racist (or the dog hater) is desperately asking for approval of his behaviour and his way of life, and for sharing his positive endeavour to promote this behaviour and way of life.

Summarizing, the use of a derogatory term in a normal speech act gives the act a new feature, besides promoting discrimination or subordination: it is a call for joint responsibility that commits co-conversationalists to *participate* in the actual subordination and deprecation of the individuals or classes defined with a pejorative. Therefore, the use of pejoratives is not just a question of informational content, or of tacit presuppositions, but it is promotion and legitimization of that content through tacit joint acceptance. In accepting a presuppositional content we ourselves turn to be *promoters of* that content, and not only making *as if* we believe it.

This seems to me a fairly acceptable rendering of the central core of Eva's analysis concerning the relationship between assertion and endorsement:

> I may refuse to endorse an assertion because its wording suggests a picture of reality that I do not share. (Picardi 2006: 62)

The central point of the refusal to endorse is rejecting the call for joint responsibility and leaving the racist alone. And probably, under this "illocutionary" view, we are allowed to say something stronger: the use of words not only "suggests" a picture of reality, but also actually "promotes" it.

7 Summary

The connection between the speech act and the set of inferences connected (either because of implicatures or because of presuppositions) with the pejorative expression builds up a new challenging problem on the relationship between truth conditions and assertibility conditions, and this seems to be the most relevant suggestion left by Eva's paper. Rejecting to endorse an assertion containing derogatory terms aims both at preventing the derivations of other assertions whose content would entail what we regard as false, and at preventing the promotion of what we considered wrong attitudes towards the object of contempt.[7]

Notes

1. A short summary of different perspective is given in Bianchi (2014), Bianchi (2015) (also with reference to experimental approaches) and Cepollaro (2015).
2. The main point is always to antagonize the *reductio* of the phenomenon of tone to the subjective alone, as Picardi (2007: 500) insists: "Tone is as much as conventional and objective feature of word meaning as sense is, and Frege erred in confining it to the realm of psychological association".
3. It is in his "Retrospective Epilog" included in Grice (1989), that Grice speaks of "Lower order" and "Higher order" Speech acts.
4. Thanks to Laura Caponetto for suggesting this connection.
5. Apparently we do not need a pejorative for an act of subordination, as in Langton's example "Blacks are not permitted to vote".
6. Or, at least, sharing the presuppositions connected with the derogatory terms. Some people may be unaware of the derogatory aspects of a term, either by not having another "politically correct" term or just by not knowing the derogatory aspect of a term in a context of a community. Travelling abroad may put people at risk of being considered either racist or simply unpolite just by ignorance (See also Penco 2017b).

7. I would like to thank Filippo Domaneschi for his suggestions on an early draft of this paper and Paolo Leonardi for his careful reading and further suggestions, which, unfortunately, I feel to have been unable to follow properly. A special thank to Laura Caponetto for pointing out some mistakes and suggesting repair.

References

Bach, K. 1999. The Myth of Conventional Implicatures. *Linguistics and Philosophy* 229: 327–366.

Bianchi, C. 2014. The Speech Acts Account of Derogatory Epithets: Some Critical Notes. *Liber Amicorum Pascal Engel*, ed. J. Dutant, D. Fassio, and A. Meylan. Geneva: Université de Genève.

Bianchi, C. 2015. A Plea for an Experimental Approach on Slurs, *Language Sciences* 52: 241–250.

Cepollaro, B. 2015. In Defense of a Presuppositional Account of Slurs. *Language Sciences* 52: 36–45.

Domaneschi, F., E. Carrea, C. Penco, and A. Greco. 2014. An Experimental Approach to Attitudes Towards Propositions. *Pragmatics and Cognition*, 291–308.

Dummett, M. 1981. *Frege, Philosophy of Language*, 2nd ed. London: Duckworth.

Dummett, M. 2007. Reply to Eva Picardi. In *The Philosophy of Michael Dummett* (LLP), ed. Schilpp and Hahn. LaSalle: Open Court.

Frege, G. 1879. *Begriffsschrift*. Halle: L. Nebert.

Frege, G. 1892. Über Sinn un Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50.

Frege, G. 1897. Logik. In *Nachgelassene Schriften*, ed. H. Hermes et al., Hamburg: Meiner. 1969; English Translation by P. Long and R. White, *Posthumous Writings*, Oxford, Blackwell, 1979.

García-Carpintero, M. 2015. Contexts as Shared Commitments. *Frontiers of Psychology*. https://doi.org/10.3389/fpsyg.2015.01932.

García-Carpintero, M. 2017. Pejoratives, Contexts and Presuppositions. *Lecture Notes in Artificial Intelligence* 10257: 15–24.

Grice, P.H. 1967. Logic and Conversation. In Chapter 2 of Grice 1989, *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grice, P.H. 1989. Retrospective Epilogue. In Grice (1989) *Studies in the Way of Words*, 339–385. Cambridge, MA: Harvard University Press.

Hom, C. 2008. The Semantics of Racial Epithets. *Journal of Philosophy* 105: 416–440.

Hom, C. 2010. Pejoratives. *Philosophy Compass* 5 (2): 164–185.

Hom, C. 2012. The Puzzle of Pejoratives. *Philosophical Studies* 159: 383–405.

Hornsby, J. 2003. Free Speech and Hate Speech: Language and Rights. In *Normatività, fatti, valori, Macerata*, ed. R. Egidi, M. Dell'Utri, and M. De Caro, 297–310. Macerata: Quodlibet.

Kaplan, D. 1999. The Meaning of Ouch and Oops. Exploration in the Theory of Meaning as Use, Unpublished.

Kissine, M. 2013. *From Utterances to Speech Acts*. Cambridge: Cambridge University Press.

Langton, R. 2012. Beyond Belief: Pragmatics in Hate Speech and Pornography. In *What Speech Does*, ed. McGowan and Maitra. Oxford: Oxford University Press.

Langton, R. 2017. "Blocking as Counter-Speech" paper presented at *ECAP9, Conference of the European Society for Analytic Philosophy* (München 2017). In The paper is forthcoming, ed. D. Harris, D. Fogal, and M. Moss, *New Work on Speech Acts*. New York: Oxford University Press.

Macià, Josep. 2014. "Expressive Meaning and Presupposition", handout for a talk at the *Names, Demonstratives, and Expressives* conference in Gargnano, Italy.

Maitra, I. 2012. Subordinating Speech. In *Speech and Harm: Controversies Over Free Speech*, ed. Ishani Maitra and Mary K. McGowan, 94–120. Oxford: Oxford University Press.

McGowan, M.K. 2004. Conversational Exercitives: Something Else We Do With Our Words. *Linguistics and Philosophy* 27 (1): 93–111.

McGowan, M.K. 2009. Oppressive Speech. *Australasian Journal of Philosophy* 87 (3): 389–407.

Penco, C. 2017a. Donnellan's Misdescriptions and Loose Talk. In *Reference and Representation in language and Thought*, ed. K. Korta and M. Ponte. Oxford: Oxford University Press.

Penco, C. 2017b. Prejudice and Presupposition in Offensive Language. *Nordicum-Mediterraneum* 12 (3).

Picardi, E. 2001. Compositionality. In *Grice's Heritage*, ed. G. Cosenza, 52–72. Turnhout: Brepols.

Picardi, E. 2006. Colouring, Multiple Propositions and Assertoric Content. *Grazer Philosophische Studien* 72: 49–72.

Picardi, E. 2007. On Sense, Tone, and Accompanying Thoughts. In *The Philosophy of Michael Dummett* (LLP), ed. Schilpp and Hahn. LaSalle: Open Court.

Popa-Wyatt, M., and Wyatt, J.L. 2017. Slurs, Roles and Power. *Philosophical Studies*. https://doi.org/10.1007/s11098-017-0986-2.

Potts, C. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Sainsbury, M. 2001. Two Ways to Smoke a Cigarette. *Ratio* 14: 386–406.

Schlenker, P. 2007. The Expressive Dimensions. *Theoretical Linguistics* 33 (2): 237–245.

Stalnaker, R. 2002. Common Ground. *Linguistics and Philosophy* 25: 701–721. https://doi.org/10.1023/A:1020867916902.

Williamson, T. 2003. Understanding and Inference. *Proceedings of the Aristotelian Society*: 249–293.

Williamson, T. 2009. Reference, Inference and the Semantics of Pejoratives. In *The Philosophy of David Kaplan*, ed. J. Almog and P. Leonardi, 137–158. Oxford: Oxford University Press.

Fregean Presentationalism

Elisabetta Sacchi

My starting point in this paper is a couple of theses that have informed the debate in the philosophy of mind in the last decades, namely: content externalism (CE) and phenomenological internalism (PI). These theses concern the individuation conditions of two kinds of properties of mental states: *representational properties* (the properties in virtue of which a state is about something and that account for its content) and *phenomenal properties* (the properties in virtue of which a state has a phenomenal character and that account for what-it-is-like for the subject

My reflections on Frege owe a deep intellectual debt to Eva Picardi. I hasten to say, however, that the position I here qualify as "Fregean Presentationalism" is very different both in the letter and in the spirit from the Fregean theory I learned to appreciate and estimate thanks to Eva's works (Picardi 1994, 1996, 2007). What comes out from my picture is a sort of hybrid figure in which some traits of the real Frege, in particular his concern with the notion of *mode of presentation*, are combined with traits that typically belong to people in the phenomenological tradition. Even though I suspect that this bizarre operation would not have met Eva's approval, I hope not to have created a philosophical monster. At least not in her eyes.

E. Sacchi (✉)
Vita-Salute San Raffaele University, Milan, Italy
e-mail: sacchi.elisabetta@hsr.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_11

of the state to be in it). CE claims that representational properties are to be individuated relationally in terms of worldly environmental features, whereas PI claims that phenomenal properties are to be individuated only by reference to the subject's intrinsic, non-relational features. Even though this combination of claims is not without opponents, it is widely taken to be the correct position as regards the metaphysics of mental properties, because it is the one that best conforms to our pre-theoretical intuitions on the nature of those properties.

In this chapter I shall not argue either for (CE)[1] or for (PI).[2] What I shall do instead is to consider which way, if any, of conceiving the relationship between representational and phenomenal properties makes the conjunction between (CE) and (PI) tenable. My main claim as regards this issue is that the conjunction is tenable only within an account which treats the two kinds of properties as distinct and irreducible to each other. After having presented and ruled out in Sect. 1. some accounts that do not satisfy this requirement, in Sect. 2. I shall put forward my suggestion on how to provide an account that complies with such requirement. According to my proposal, that I shall label (for reasons that will become clear hereinafter) "Fregean Presentationalism", *phenomenal properties are manners of presentation of representational properties.* Finally, in Sect. 3. I shall consider a critical point raised by Ellis (2010) who, on the basis of considerations concerning the kind of introspective access people have to the phenomenal character of their own conscious states, aims at showing that the conjunction of (CE) and (PI) is untenable whatever one's favorite metaphysical model of phenomenal character happens to be. I shall address this critical point by showing that it presupposes a picture of inner awareness that is far from being mandatory. In particular, I shall claim that the suggested picture is completely at odds with the kind of metaphysics of phenomenal properties that Fregean Presentationalism recommends.

1 What Does Challenge the Tenability of the Conjunction

In this section I shall present two ways of accounting for the relationship between representational and phenomenal properties that, however appealing, are unsuited to make the conjunction of (CE) and (PI) tenable. Both ways share the claim that mental phenomena constitute

a homogeneous domain unified under one single kind of property; this property is taken to be the most basic one from both a metaphysical and an explanatory point of view. According to one view, that we shall call "strong representationalism" (SR), the property in question is *intentionality* (understood as the property of mental states that makes them *represent* something). According to (SR) the representational properties of mental states do not depend (either ontologically, or explanatorily) on phenomenal properties; by contrast, the latter are claimed to depend on and be reducible to the former. The alternative view, that we shall label "strong phenomenalism" (SP), claims instead that the most basic mental property is *phenomenality* (the property for a mental state to have a phenomenal character). The order of dependence among the two properties according to (SP) is just the opposite of that outlined by (SR): phenomenality comes first and intentionality is either constitutively grounded in it (as it is the case with people endorsing the "phenomenal intentionality" thesis[3]) or derived from the kind of intentionality which is so grounded.[4]

Both (SR) and (SP) provide clear answers to the issue of the relationship between representational and phenomenal properties. This answer is cashed out in terms of identity claims (IC), namely and respectively:

(SRIC) phenomenal properties are identical with (some kind of) representational properties
(SPIC) representational properties are identical with (some kind of) phenomenal properties.

The fact that neither position is able to account for [(CE) & (PI)] can be easily shown by presenting a trilemma that each position faces as soon as its respective identity claim is combined with two propositions that correspond to CE and PI. The trilemma of the strong representationalist view is the following:

1. Phenomenal properties are identical with (some kind of) representational properties (SRIC)
2. Representational properties are wide (CE)
3. Phenomenal properties are narrow (PI)

As a matter of fact, these three propositions are jointly inconsistent, no matter how intuitively appealing they can look when taken on their own. If phenomenal properties are a proper subset of representational

properties, it follows that they should be individuated in the same way as the other properties in the set. Therefore, if one claims that the correct individuation to provide for the properties in the set is wide, one has to reject phenomenological internalism in favor of phenomenological externalism.[5] By contrast, if one feels unease with this upshot because one cannot abandon the idea that phenomenology is narrow,[6] one has no option but to reject (CE) and embrace content internalism.[7]

The trilemma that strong phenomenalism faces (that differs from the previous only as regards the first proposition—i.e. the identity claim endorsed by this position) is the following:

1*. Representational properties are identical with (some kind of) phenomenal properties (SPIC)
2. Representational properties are wide (CE)
3. Phenomenal properties are narrow (PI)

The considerations made as regards the representationalist version of the trilemma apply *mutatis mutandis* to the present version as well. It is worth noting, however, that even though there actually are, as in the previous case, two possible ways out of the inconsistent triad, in general those who endorse (SPIC) tend to reject (2) and stick to (3).[8]

It is worth stressing that both (SR) and (SP) can defend the conjunction of (PI) with a restricted version of (CE), let us label it (CE)*, that says that *at least some* representational properties are wide. This result can be achieved by denying premise (2) of the inconsistent triads not by affirming the contrary proposition (no representational property is wide), but the contradictory one (not all representational properties are wide). This move brings about a "dual picture" that takes the content of a mental state to be composed by two components, a narrow component (shared by a subject and any of her intrinsic duplicates, notwithstanding how their respective physical/social environments happen to be) and a wide component (a component that is not shared). Letting aside all the problems that the accounts of narrow content notoriously meet, it goes without saying that this is not a way in which, strictly speaking, the conjunction of (CE) and (PI) could be said to be vindicated. That's why I shall no more consider this theoretical option in the foregoing. The question I shall now address is whether there is some other way of accounting for the relationship between representational and

phenomenological properties that is able to save both (CE) and (PI) in their strong readings.

If this question had been put to any philosopher in the analytic tradition in the not so remote past, she would hardly have seen any difficulty in providing a positive answer to it. I am thinking in particular of those philosophers, like G. Ryle, C. Lewis and W. Sellars for example, who adhered to what we now call (starting from C. Lewis who first introduced the label) the "two-separate-realms" conception of the mind. This picture takes the mind to be constituted by two different and irreducible domains of phenomena: the purely representational domain (the domain of states such as beliefs, desires) and the purely qualitative domain (to which phenomena such as moods, bodily feelings, sensations, belong). The former is the domain of propositional attitudes, of mental states endowed with content (i.e. that have intentionality); the latter is the domain of states that feel like something to the subject who entertains them (i.e. that have phenomenal character). In this framework, phenomenality is taken to pertain only to states of the latter kind, whereas intentionality only to states of the former kind. According to this conception, the representational and the phenomenal properties belong to two distinct and independent kinds: a state can exemplify one kind without exemplifying the other and when the two properties happen to be co-instantiated they are taken to be logically and ontologically independent from each other. As things stand, within this picture there is no problem in holding that the two kinds of properties have different individuation conditions. And yet, even though the two-separate-realms conception allows one to provide a positive answer to our initial question, the picture of the mind that it promotes has very little to recommend itself. As a matter of fact, the idea that the mind is a dishomogenous field of phenomena deprived of any unifying factor constitutes an unbearable obstacle to the attempt to achieve a thorough understanding of the mental domain. Even granting that the representational and the phenomenal properties could eventually differ in their extension, it is uncontroversial that there are cases in which they are co-instantiated. A case in point is provided by perceptual states. The advocate of the two-separate-realms conception of the mind allows for such states to exemplify both properties, but she conceives them as unrelated. And yet, the idea that the phenomenal and the representational aspects of our mental life are so disconnected is hard to swallow.[9] Given the several drawbacks and inadequacies of such a position, if it turned out that it is the only available option, that could

be taken as a *reductio* of the claim that the conjunction is tenable. So the question to consider is whether there is some alternative to the strong reductive versions, different from the "two-separate-realms conception", that could enable one to save both (CE) and (PI) in their strong readings. In order to pass muster, one such alternative should satisfy at least the two following requirements: (i) reject the identity claims of both (SR) and (SP) and (ii) provide a different account of the relationship between the representational and the phenomenal properties.[10] For if no such account is provided, the resulting picture will inevitably inherit most of the problems affecting the two-separate-realms conception. Well, is there any position that passes muster?

Most people would say that there is a simple answer to such a question: all one has to do is to abandon the identity claims in favor of some claim of supervenience and that's it. I disagree on this point. In my view, just embracing some supervenience claim is not enough, because, as Kim (1993) has vigorously stressed, all that supervenience can account for is the co-variation of one kind of property with another kind of property, without settling the further question of what grounds the co-variation itself. Co-variation needs to be based on some relation of dependence between the kinds of properties involved and this relation, in turn, has to be grounded in a metaphysical picture of the natures of those properties. In the next section I shall put forward my positive proposal on this issue.

2 Fregean Presentationalism

According to my proposal phenomenal properties and representational properties are two distinct and irreducible kinds of mental properties (so I reject both SRIC and SPIC). As a consequence, the individuation of each one has no bearing on the individuation of the other. While being distinct and irreducible, I claim that they stand to each other in a peculiar relationship which is grounded in their very nature. Phenomenal properties in my picture are *manners of presentation* to the subject of representational properties, where a manner of presentation of a representational property is the way in which a given representational property is (phenomenally) presented to the subject of the experience or, to put it in other words, the way in which a representational property is experientially given to the subject. That is why an appropriate label to use for them would be "presentational properties".

That conscious mental states have not only a representational side, but also a presentational one (or, what amounts to the same thing, the fact that what is represented in such states is always experientially presented to the subject undergoing the experience) is a point which has been widely stressed within the phenomenological tradition, starting from Brentano who underlined the reflexive character of conscious mental states. Within the analytic tradition, this point has been vigorously emphasized, among others, by C. McGinn (1988) and, in particular, by his talk of the "Janus-faced" character of conscious intentionality having both an "outward-looking face" (an objective side) and an "inward-looking face" (a subjective side).

What I am here adding to this traditional picture is a sort of Fregean reading of the two faces/sides that constitute the structure of conscious mental states. According to this reading, in any conscious mental state something is *represented* (the objective side) and what is represented (what the state is directed to) is always *presented* to the conscious subject in a given manner (the subjective side).[11] A manner of presentation is what makes a *representatum* (a represented object/property/state of affairs) experientially available to the experiencing subject from his/her point of view. Phenomenal properties can therefore be taken as the experiential counterparts of Fregean modes of presentation in so far as they play, in the subject's experience, an experiential presentational role. While modes of presentation play a *cognitive* presentational role in virtue of being *ways of thinking* (where a way of thinking of something is a way of having individuative knowledge of it),[12] manners of presentation play an *experiential* presentational role in virtue of being *ways of experiencing* something.

The idea that phenomenal properties play a presentational role in our experience and that this role is constitutive of their metaphysical nature, features very clearly in Chalmers's (2004) version of representationalism that he accordingly qualifies as Fregean. Fregean Representationalism, according to Chalmers, is the claim that phenomenal properties are identical to certain *Fregean representational properties*, where a Fregean representational property is characterized as the property of having (in a certain way) a certain Fregean content. The content qualifies as Fregean because it involves a mode of presentation whose role is to pick up what is represented in a given experiential state. Take phenomenal redness as an example of a phenomenal property. According to Chalmers's picture, phenomenal redness is identical to the representational property

of representing in a given way (visually phenomenally) a given property (red, say, for people leaving in our environment and with our kind of perceptual system) under the mode of presentation *the property that normally causes experiences of phenomenal redness*. Chalmers uses the label "manners of representation" for the various ways in which a given content can be entertained,[13] and he distinguishes *manners* of representation from *modes* of presentation. Whereas the former are psychological entities (namely: psychological features of individuals), the latter are abstract entities to which psychological states may be related, by having them as their content.[14]

Fregean Presentationalism is different from Fregean Representationalism as presented by Chalmers (2004). The main line of disagreement has to do with the claim, which he endorses and I reject, that phenomenal properties belong to the content of the act. Let me expand on this point starting from the notion of content itself. I deem it important in addressing the issue as to whether phenomenal properties do belong to the content of a mental state to work with a characterization of the notions at hand, in particular the notion of *content*, that is as neutral as possible. This requirement is not satisfied in my view by a characterization, quite widespread nowadays, according to which the content of a mental episode is everything that one experiences in entertaining the mental episode, or everything that is giving to one experientially in having the experience, or everything one is aware of. Of course, if one adopts this "maximally inclusive conception of content",[15] the question we are raising turns out to be settled on mere definitional grounds. In my view this way of characterizing the notion of content should be avoided because, for a verse, it ends up trivializing an issue which is very substantial and not merely terminological and, for another, it is not sufficiently restrictive. Regarding this last point, there seem to be many things one can be aware of in having a given experience that do not belong to the content of the state. Think for example of the awareness that we have of the kind of experience (visual, tactile or otherwise) we are enjoying. This is something we are aware of in having the experience and yet we would not take this aspect of our awareness to belong to the content of the state. That's why I think it preferable in dealing with the issue at hand to work with a more traditional and less broad characterization of the notion according to which the content of a mental state is what a given mental state represents and that accounts for its correctness conditions.

To come back to my critical point, the risk which I see in putting the phenomenal properties of a state in its content is that this move tends to promote a "subjectivization" of content according to which the content would be constituted not only by objective properties (properties of the represented object), but also by subjective, mind-dependent properties (properties of the way in which something phenomenally looks or appears to the subject), making in this way impossible for distinct subjects (or even for the same subject at different times) to entertain one and the same content. A possible rejoinder on the part of our opponent could be to claim that those alleged mind-dependent properties that account for how something is given to the experiencing subject actually are objective properties of the represented object, albeit response-dependent ones (that is: dispositional properties to elicit certain responses in certain respondents under certain conditions).[16] Even granting that this move would enable one to avoid the above criticism, the suspicion still remains that it runs the risk of introducing between the mind and the world a "veil of appearance" that would seriously challenge the idea, dear to the representationalist, that our experience is "transparent", that is: it puts us directly in contact with the external world.

Having clarified the Fregean spirit of my proposal, and its connection with Chalmers's, I hasten to add some important differences, besides the ones already stressed, between Fregean modes of presentation and my manners of presentation. This will enable me to highlight one crucial difference between my way of interpreting the presentational role of phenomenal properties and Chalmers's. The crucial difference is that, while playing a presentational role in the experience, *manners of presentations are not senses* as Fregean modes of presentation are. What I mean by this is that manners of presentation, as I take them, are not features (either conceptual or non-conceptual) figuring in the content of a mental state, as instead Chalmers holds of his modes of presentation. I agree with Chalmers that phenomenal properties play a presentational role.[17] But whereas he interprets this role along Frege's lines (i.e. to play this role is for a feature to provide an identifying condition that picks up what is represented in a given state) I interpret it differently. In my view, phenomenal properties are not experiential identifying conditions that pick up a *representatum*, rather they are ways of acquaintance of represented properties. Both Chalmers's position and mine can be qualified as Fregean. In a sense Chalmers's position is more faithful to Frege's conception of modes of presentation, because he takes them to feature in the

content of the state. And yet, since the modes of presentation that figure in conscious mental states have a phenomenological-subjective nature, his position turns out to be unFregean in so far as it gets rid of the idea that contents are objective and mind-independent entities.[18] This is instead a tenet that my position retains.

The above considerations leave the following question open: If phenomenal properties are not content properties, what are they properties of? My answer to this question is that phenomenal properties belong to that other feature of mental states that, together with content, is taken to individuate them, namely: the intentional mode. The intentional mode of a state is the way in which the content of the state is entertained. In general, this is specified in terms of the kind of state involved: if the state is perceptual the content is entertained in the perceptual modality (which can be visual, auditory, tactile and so on and so forth); if it is a propositional attitude, the content is entertained in some attitudinal modality (in the belief modality, in the desire modality, and so on and so forth). The distinction between intentional mode and intentional content that I am here recruiting is familiar in the philosophical tradition even though different authors have used different labels for it.[19] What I am here suggesting is that the notion of intentional mode, as far as conscious mental states are concerned, should be broadened in such a way as to encompass not only the psychological components of the state but also its phenomenological components.[20]

Before articulating my proposal, let me try to clarify some points. What does it mean to say that the intentional mode of a conscious state has more structure than that of an unconscious state? Let us consider for example a visual perception. How does the intentional mode of a conscious seeing differ from the intentional mode of an unconscious seeing (think for example to a case of inattentional blindness)? Isn't the intentional mode the same in the two cases, namely a visual modality? What kind of properties the former presents and the latter lacks that could account for their phenomenological difference? I think that a perspicuous way to clarify this point is to make use of the idea that mental states are kind of relations to mental representations which are bearers of content. Taking this framework in the background, what I am saying is that the kind of mental representation which is instantiated in the two cases is different. The difference in question has to do with the structure of the mental representation itself. While the representation that gets mobilized in an unconscious state has only a representational component,

the one that occurs in a conscious state has also a presentational component: what is represented is presented (i) to the subject (ii) in a given manner. This manner is not exhausted by the psychological aspect of the intentional mode (i.e. that aspect of it that accounts for the kind of experiential act involved: whether it is a visual experience, an auditory experience, a tactile experience and so on and so forth). Consciously seeing a red apple differs from consciously seeing a green apple. The psychological aspect involved is the same, that of vision, but the way in which the two color properties are visually presented is different. The difference at issue is one in ways of seeing. It is in this sense that one has to understand the claim that the intentional mode of a conscious state has more structure than that of an unconscious state.

Let me now present my position in more details. One central claim of Fregean Presentationalism is that the intentional mode of a conscious state has a subjective dimension that no intentional mode of any unconscious state has. This subjective dimension in turn has (at least) two components, a *to-me component* and a *qualitative component*, that together constitute the conscious state's phenomenal character.[21] Let me provide an elucidation of these two components starting from the former. In any conscious mental state something is presented *to* a subject; the *to-me* component is precisely the first-person presentedness which accompanies any conscious occurrent state.[22] While accounting for what makes a state a phenomenally conscious state at all, this component is not responsible for a state being the phenomenally conscious state it is, because it is common to all phenomenally conscious states of a subject. What plays that role is rather the qualitative component. This component captures the way in which what is represented (the state's content) is (experientially) presented to the subject of the state.

According to Fregean Presentationalism, the phenomenal character of a certain conscious state is a matter of experiencing in a certain sort of way, or a way of experiencing, a way the experiential act is. This way of putting things goes in the direction of an *adverbialist account* along the lines indicated by Thomasson (2000). According to adverbialism in the form I wish to endorse, when I consciously experience the blue of the sky, the bluish way it is like for me to see the sky is a modification of my way of seeing; I am seeing-blue-wise and the fact that I am so seeing is what accounts for the phenomenal character of my experience.

Having clarified the main ideas behind Fregean presentationalism let me now go back to the question I raised at the beginning as to whether

it is possible to defend the conjunction of (CE) and (PI) without having to endorse some version of the two-separate-realms conception of the mind. According to my proposal, what one has to do to that end is to get rid of the metaphysical claim that the two properties are reducible to each other. By abandoning this claim (rejecting both proposition 1. and proposition 1*. of the triads I presented in the first section) one can maintain that one kind of properties (the phenomenal ones) is internalistically individuated whereas the other kind (the representational ones) is externalistically individuated. So my first tentative conclusion regarding the question I raised at the beginning is in the affirmative. As I said, this is a tentative conclusion. As it will turn out, one further step is required. If the first step had to do with metaphysical issues (issues concerning the nature of the properties involved), the second step has to do with issues in epistemology, in particular with issues concerning the kind of introspective knowledge we have of our conscious mental states. This is the topic of the next section.

3 Phenomenal Character and Introspective Accessibility

In a recent paper Ellis (Ellis 2010) has claimed that the conjunction of externalism about content and internalism about phenomenal character is untenable whatever one's favorite model of phenomenal character happens to be. If Ellis were right, it would follow that, contrary to what I have tried to do, one could not be an externalist about content and an internalist about phenomenal character; if externalism is true at all, it ought to be true of both content and phenomenal character. The only claim that in his view has to be endorsed in order for his argument to go through[23] is one that everyone is pretty willing to accept, namely that the phenomenal properties of a mental state are in principle accessible in introspection. His argument is that whoever endorses that claim is thereby compelled to endorse two theses that together entail externalism about phenomenal character. The two theses are the following: (A) *Accessibility*: for any phenomenal property of a subject's experience, if the subject introspectively attends to her experience, she can employ a phenomenal concept that refers to that property[24]; and (TI) *Twin Inability*: it is possible for a subject to have an experience with a phenomenal property that is not referred to by any phenomenal concept that the subject's (introspectively

capable) twin can employ. The notion of *phenomenal concept* occurs in both theses and it is crucial in Ellis's argument: he bases his claim that the conjunction of (CE) and (PI) is untenable on any model of phenomenal character precisely on works on phenomenal concepts. According to Ellis if one endorses content externalism, one also has to endorse externalism about phenomenal character, because externalism about content implies externalism about phenomenal concepts which, in turn, implies externalism about phenomenal character.[25]

In his paper Ellis focuses on content externalists who reject representationalism and who hold that the phenomenal properties of mental states are in principle introspectively accessible. Even though I have said nothing so far about this issue, I do endorse the claim that the phenomenal properties of our mental states are in principle introspectively accessible.[26] Thus, his argument is targeted exactly against a position such as the one I have put forward. Given that a Fregean presentationalist is precisely a philosopher of the relevant kind (she endorses content externalism, she rejects representationalism, she accepts introspective accessibility), then, if Ellis were right, a Fregean presentationalist should accept both (A) and (TI) that together, he claims, implies externalism about phenomenal character.[27,28]

A possibility which he considers is that such a philosopher could either abandon the idea that we have introspective access to the phenomenal properties of our conscious mental states or otherwise abandon the way of conceiving introspective accessibility that (A) requires and significantly re-think the very notion of introspective accessibility. I agree with him that the first horn is a desperate one, because it has very unpalatable consequences. But the second horn seems to me exactly the right move to take in order to preserve the conjunction of CE and PI. But what would such a re-thinking consist in?

Let us come back to Ellis's first premise (A). Curiously, Ellis doesn't spend much time in justifying why one who is willing to endorse the idea of introspective accessibility should thereby accept (A). As a matter of fact, one could be resistant in introducing phenomenal concepts in one's ontology or in allowing them to play such a substantive role in inner awareness.[29,30] All that Ellis provides in support of (A) is the following consideration: "it will suffice for our purpose to note that when a subject is introspectively aware of a phenomenal quality, the awareness is conceptual; it involves the employment of concepts. This idea is orthodox now that most philosophers have abandoned so called 'inner sense'

or 'perception-object' models of introspection, on which we observe or perceive our own mental states and their properties" (Ellis 2010: 278). Ellis's argument takes for granted that the only alternative to a perception-object model of introspection is a thought-object model. The two models are variants of what can be called an *act-object model of introspection* according to which the kind of awareness involved in introspection is of the same kind as that involved in exteroception, that is a kind of object-awareness, a representational relation whose *relata*, however different in the two cases (the phenomenal character of a conscious mental state in one case and an external worldly item in the other), play the same role, namely: the intentional object role. In my view, the mistake of Ellis's argument is to assume that the act-object model of introspection is one that everyone should endorse, whatever one's favorite metaphysics of phenomenal character turns out to be. As a matter of fact, if the act-object model of introspection is perfectly at home and consonant with a representational account of phenomenal character, it is not at all consonant with an adverbial account such as the one I have presented. *Contra* Ellis, I think that not only an adverbialist is not compelled to endorse such a model, but rather she ought not to, because such a model is patently in conflict with the way in which such a theorist conceives the metaphysics of phenomenal character. An adverbialist metaphysics of phenomenal character is at odds with an act-object epistemological model of inner awareness.[31] If the phenomenal character of an experience is the way in which what is represented is (experientially) presented to the subject, if in other words, it is a way of presentation of something to the very subject undergoing the experience, then the phenomenal character has to be directly and unmediatedly given to the subject of the experience, but not as an object to be perceived or to be picked up by some concept, be it phenomenal or not, but rather as a way of experiencing that makes the subject aware of what is represented in her experience.[32] In this way, by ruling out the idea that inner awareness is achieved through the mediating role of phenomenal concepts, Ellis's argument can no more go through.[33] *Contra* Ellis, I therefore claim that the conjunction of (CE) and (PI) is tenable. I agree with him that the mere rejection of strong representationalism is not enough in order to preserve the conjunction. What one has to do, if I am right, is rather the following: (i) reject both strong representationalism (SR) and strong phenomenalism (SP); (ii) account for the relationship between the representational and the phenomenal properties in such a way as

not to make them dependent on each other for their individuation; and (iii) reject an act-object model of the kind of introspective access we have of phenomenal properties. I think that the position I have outlined satisfies these requirements. Therefore, I conclude by saying that if Fregean presentationalism is a legitimate position, then it ultimately turns out to be possible to account for phenomenal character so as to save our most rooted intuitions about how content and phenomenology have to be individuated.

Notes

1. One can find three main groups of arguments in support of content externalism: (i) Putnam's Twin-Earth arguments; (ii) Burge's arguments concerning the social individuation of concepts; and (iii) McDowell's and Davidson's arguments motivated by normativity claims.
2. Typically, the phenomenological internalist thesis is merely asserted, because it is considered to be so obvious as not to require any justification. An exception is Pautz (2006).
3. For a recent overview of the debate on phenomenal intentionality see Bourget, David, and Mendelovici, Angela, "Phenomenal Intentionality", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2017/entries/phenomenal-intentionality/.
4. For a presentation of this theoretical stance see Kriegel's introduction in Kriegel (2013) where the author illustrates the various ways in which the "constitutive grounding claim" and the "derivative claim" are understood by people adhering to what is labelled the "Phenomenal intentionality research project".
5. This position is exemplified within the representationalist camp by Dretske (1995), Lycan (1996, 2001), Tye (1995). Of course, phenomenal externalism is not confined to representationalism. An example of an externalist and yet non-representationalist position is Martin's (2002, 2006).
6. For a critical discussion of phenomenological externalism see Sacchi and Voltolini (2017).
7. Among those who reject (2) there are Chalmers (2004), Kriegel (2002), Levine (2003), Thau (2002).
8. This position is variously endorsed by Loar (2003), Horgan and Tienson (2002), Horgan et al. (2004), Georgalis (2006), Kriegel (2007, 2011), Farkas (2008).
9. One drawback of this position is its inability to account for the intuition that phenomenal duplicates are similar also under many representational respects.

10. In general, people who reject the identity claims and defend the conjunction—as for example Burge (1979, 2003), Shoemaker (1994)—either do not worry at all or only very little about the second requirement.
11. The reason why I am using "manners of presentation" instead of "modes of presentation" will become clear in the foregoing.
12. This is the standard characterization of senses within the neo-Fregean tradition inaugurated by Evans (1982).
13. His notion, as he himself acknowledges, is strictly connected with the notion of the "intentional mode" of the mental act/state.
14. Modes of presentation, according to Chalmers, qualify as contents (or constituents of them) because they have built-in conditions of satisfaction (they specify the conditions that something, an object or a property, has to satisfy in order to be the object/property represented in a given experience).
15. See Montague (2010) for this characterization of the notion of content.
16. For this characterization of the notion of *response-dependent properties* see Kriegel (2009).
17. I articulated this idea in Sacchi (2011).
18. That no subjective aspect can occur in the content of a thought is central in Frege's anti-psychologism, as has been vigorously stressed by Picardi (1996). Frege's anti-psychologism has been attacked by many people within the so called cognitive turn in analytic philosophy. I side with those people who claim that the neat divide that Frege drew between the subjective "realm" and the objective "realm" is problematic and that a better picture of mental activities and their products should encompass Frege's strictures. I have defended this claim in Sacchi (2007). Nonetheless, I am strongly suspicious of those recent moves that aim at "phenomenologizing" the content of mental states. One such move has been recently advanced by Pitt (2009) in his plea for what he labels "intentional psychologism". Intentional "psychologism", as Pitt qualifies it, is the thesis that intentional mental content is phenomenological (in the strong sense that the phenomenology of mental states *constitutes* mental content—i.e. thoughts with the same phenomenology have the same intentional content). This is a form of psychologism, Pitt stresses, because it treats the contents of our thoughts (supposedly abstract objects) as some kind of mental objects, namely: cognitive phenomenal objects.
19. Meinong, Twardowski and Husserl distinguish among the *quality* and the *matter* of an act. Smith and McIntyre talk of *thetic character* and *noematic Sinn*. Our suggested way of framing the distinction is standard in the debate. See e.g. Searle (1983), Crane (2001).
20. The idea that the phenomenal character of a mental state can be explained in terms of its intentional mode, understood in a suitably widened way, can be traced back to D.W. Smith (1986, 1989, 2005).

21. My distinction matches the one drawn by Kriegel (2009) between subjective character (which he labels "for-me component") and qualitative character as two aspects or components of phenomenal character.
22. On the notion of presentedness or phenomenal presence see Frey (2013) in which the author clarifies how this notion is fundamental for understanding the peculiarity of experiential or phenomenal intentionality.
23. Actually, as it will become clear, one more claim, far less trivial, is implied, namely that introspective accessibility or inner awareness involves the employment of phenomenal concepts.
24. By *phenomenal concepts* he means "those primary concepts through which we think about phenomenal properties" (Ellis 2010: 289).
25. As he clarifies (see footnote 25 on p. 289) his treatment of these issues strongly resembles Dretske's (1995, Chapter 5). But the way in which externalism about phenomenal concepts is derived from externalism about content and the way in which externalism about phenomenal character is derived from externalism about phenomenal concepts is different in the two authors. While in Dretske both implications follow from assuming the representationalist model, in Ellis this is not so, because he wants his argument to be accepted also by those who resist representationalism and think, wrongly in his view, that this rejection is sufficient in order to maintain the conjunction of (CE) and (PI).
26. As it will become clear, even though I endorse this claim I disagree with Ellis's way of accounting for it (namely with the idea that in order to account for it one has to endorse thesis (**A**)).
27. This is how Ellis explains the implication: "If Sally has an experience with a phenomenal quality Q, and none of the phenomenal concepts that Twin Sally can employ refers to Q, then Twin Sally's experience cannot have the phenomenal quality Q. For if it did, then by Accessibility, Twin Sally could employ a phenomenal concept that referred to Q" (Ellis 2010: 277).
28. Actually, the structure of Ellis's argument is more complex than I have sketched. He claims that a content externalist is committed to (TI) because he must accept two other premises that together lead to (TI). These premises are: (i) externalism about phenomenal concepts (EPC): The phenomenal concepts that a subject is able to employ could be substantially different from the phenomenal concepts that her physical twin is able to employ; and (ii) conceptual co-reference (CR): Physical duplicates could not refer to the same phenomenal quality with different phenomenal concepts. So he claims that a content externalist must accept (CR) and thus (TI) if she accepts (EPC). And since (A) and (TI) entail externalism about phenomenal character she must accept that too.

29. On this point see e.g. Prinz (2007), Tye (2009) who present accounts of phenomenal knowledge that do not depend on phenomenal concepts because they adopt an eliminativist view on them.
30. According to Ellis, one could not hold that phenomenal character is introspectively accessible and adopt an eliminativist account of phenomenal concepts. For, he claims, "If introspection involves the employment of phenomenal concepts, then to deny that we have phenomenal concepts would render one unable to explain the accessibility of phenomenal qualities" (Ellis 2010: 295). The problem, of course, is that the antecedent of the conditional (that introspection involves the employment of phenomenal concepts) is not something that everyone is willing to accept.
31. Thomasson (2000), Zahavi (2004).
32. One could object that the "directedness" of the access is not challenged by the idea that the phenomenal properties of our experience are picked up by phenomenal concepts, because on nearly all accounts, phenomenal concepts are taken to refer directly (see, e.g. Loar 1997, Papineau 2002, Chalmers 2003). The way in which the objector conceives the notion of "directedness" is that there is no intermediate descriptive reference-fixing mode of presentation between a given phenomenal concept and the phenomenal property that is accessed through it and made available as an object of inner awareness. In my view, this is not a sufficiently radical reading of the directedness of our access to those properties. We are directly aware of them (if we attentively attend to them), because they are self-presenting. And this follows from their being "manners of presentation". The point here is analogous to what holds for Fregean senses. What plays the presentational role does not have in its turn to be presented, because what does the presenting is itself, on pain of circularity, self-presenting.
33. This is so because in that argument the externalist conclusion about phenomenal character is reached through a substantive use of the externalist thesis about phenomenal concepts.

References

Bourget, D., and Mendelovici, A. 2016. Phenomenal Intentionality. In *The Stanford Encyclopedia of Philosophy* (Fall 2016, Edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/spr2017/entries/phenomenal-intentionality/.

Burge, T. 1979. Individualism and the Mental. In *Midwest Studies in Philosophy*, vol. 4, ed. P. French, T. Uheling, and H. Wettstein. Minneapolis: University of Minnesota Press.

Burge, T. 2003. Perceptual Entitlement. *Philosophy and Phenomenological Research* 67: 503–548.

Chalmers, D. 2003. The Content and Epistemology of Phenomenal Belief. In *Consciousness: New Philosophical Perspectives*, ed. Q. Smith and A. Jokic, 220–227. New York: Oxford University Press.

Chalmers, D. 2004. The Representational Character of Experience. In *The Future for Philosophy*, ed. B. Leiter. Oxford: Oxford University Press. Rpt. in Chalmers 2010. *The Character of Consciousness*, 339–379. Oxford: Oxford University Press.

Crane, T. 2001. *Elements of Mind*. Oxford: Oxford University Press.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Ellis, J. 2010. Phenomenal Character, Phenomenal Concepts, and Externalism. *Philosophical Studies* 147: 273–299.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Farkas, K. 2008. Phenomenal Intentionality Without Compromise. *The Monist* 71 (2): 273–293.

Frey, C. 2013. Phenomenal Presence. In *Phenomenal Intentionality*, ed. U. Kriegel. Oxford: Oxford University Press.

Georgalis, N. 2006. *The Primacy of the Subjective: Foundations for a Unified Theory of Mind and Language*. Cambridge, MA: MIT Press.

Horgan, T., and J. Tienson. 2002. The Intentionality of Phenomenology and the Phenomenology of Intentionality. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. D. Chalmers. Oxford: Oxford University Press.

Horgan, T., J. Tienson, and G. Graham. 2004. Phenomenal Intentionality and the Brain in a Vat. In *The Externalist Challenge*, ed. R. Schantz. Amsterdam: Walter de Gruyter.

Kim, J. (ed.). 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kriegel, U. 2002. Phenomenal Content. *Erkenntnis* 57: 175–198.

Kriegel, U. 2007. Intentional Inexistence and Phenomenal Intentionality. *Philosophical Perspectives* 21: 307–340.

Kriegel, U. 2009. *Subjective Consciousness*. Oxford: Oxford University Press.

Kriegel, U. 2011. *The Sources of Intentionality*. Oxford and New York: Oxford University Press.

Kriegel, U. (ed.). 2013. *Phenomenal Intentionality*. Oxford: Oxford University Press.

Levine, J. 2003. Experience and Representation. In *Consciousness: New Essays*, ed. Q. Smith and A. Jokic. Oxford: Oxford University Press.

Loar, B. 1997. Phenomenal States II. In *The Nature of Consciousness: Philosophical Debates*, ed. N. Block, O. Flanagan, and G. Güzeldere. Cambridge, MA: MIT Press.

Loar, B. 2003. Phenomenal Intentionality as the Basis of Mental Content. In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, ed. M. Hahn and B. Ramberg, 229–258. Cambridge, MA: MIT Press.

Lycan, W.G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
Lycan, W.G. 2001. The Case for Phenomenological Externalism. *Philosophical Perspectives* 15: 17–35.
Martin, M.G.F. 2002. Particular Thoughts & Singular Thought. In *Logic, Thought, and Language*, ed. A. O'Hear, 173–214. Cambridge: Cambridge University Press.
Martin, M.G.F. 2006. On Being Alienated. In *Perceptual Experience*, ed. T.S. Gendler and J. Hawthorne, 356–410. Oxford: Oxford University Press.
McGinn, C. 1988. Consciousness and Content. In *The Nature of Consciousness*, ed. N. Block et al., 295–307. Cambridge MA: MIT Press.
Montague, M. 2010. Recent Work on Intentionality. *Analysis Reviews* 70 (4): 765–782.
Papineau, D. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.
Pautz, A. 2006. Sensory Awareness Is Not a Wide Physical Relation: An Empirical Argument Against Externalist Intentionalism. *Noûs* 40 (2): 205–240.
Picardi, E. 1994. *La chimica dei concetti*. Bologna: Il Mulino.
Picardi, E. 1996. Frege's Anti-psychologism. In *Frege, Importance and Legacy*, ed. M. Schirn. Berlin and New York: De Gruyter.
Picardi, E. 2007. On Sense, Tone, and Accompanying Thoughts. In *The Philosophy of Michael Dummett*, ed. R.E. Auxier and L.E. Hahn, 491–520. LaSalle: Open Court.
Pitt, D. 2009. Intentional Psychologism. *Philosophical Studies* 146 (1): 117–138.
Prinz, J. 2007. Mental Pointing: Phenomenal Knowledge Without Concepts. *Journal of Consciousness Studies* 14: 184–211.
Sacchi, E. 2007. Thought and Thinking: The Ontological Ground of Intentionality. In *Explaining the Mental. Naturalist and Non-Naturalist Approaches to Mental Acts and Processes*, ed. M. Beaney, C. Penco, and M. Vignolo, 135–152. Cambridge: Cambridge Scholar Press.
Sacchi, E. 2011. Aspectuality and Phenomenal Character: Towards a Phenomenologically Oriented Account of the Phenomenon of Aspectuality in Propositional Attitudes. *Phenomenology and Mind* 1: 58–71.
Sacchi, E., and A. Voltolini. 2017. Against Phenomenological Externalism. *Critica* 49 (145): 25–48.
Searle, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.
Shoemaker, S. 1994. Phenomenal Character. *Noûs* 28: 21–38.
Smith, D.W. 1986. The Structure of (Self-)Consciousness. *Topoi* 5: 149–156.
Smith, D.W. 1989. *The Circle of Acquaintance*. Dordrecht: Kluver Academic Publishers.
Smith, D.W. 2005. Consciousness with Reflexive Content. In *Phenomenology and Philosophy of Mind*, ed. D.W. Smith and A. Thomasson. Oxford: Clarendon Press.

Thau, M. 2002. *Consciousness and Cognition*. Oxford: Oxford University Press.
Thomasson, A.L. 2000. After Brentano: A One-Level Theory of Consciousness. *European Journal of Philosophy* 8: 190–209.
Tye, M. 1995. *Ten Problems of Consciousness.* Cambridge, MA: MIT Press.
Tye, M. 2009. *Consciousness Revisited: Materialism without Phenomenal Concepts.* Cambridge, MA: MIT Press.
Zahavi, D. 2004. Back to Brentano? *Journal of Consciousness Studies* 11: 66–87.

PART II

Themes from Davidson

Agency Without Rationality

Lisa Bortolotti

1 Introduction

There are many ways in which Eva Picardi's teaching and research have influenced the development of my ideas, and my engagement with philosophy as a subject and as a way of life.[1] In this chapter, I want to revisit one theme that obviously bears witness to Eva's influence, the role of rationality in interpretation, and in particular the plausibility of a rationality constraint on the ascription of intentional states and on intentional agency. Philosophers have long been interested in whether intentional agents need to be rational (to some extent, and for some given notion of rationality). I think it is fair to say that, following Donald Davidson's writings on radical interpretation and belief ascription, and Daniel Dennett's on the intentional system theory, the consensus is that some notion of rationality is a condition without which humans could not be ascribed intentional states or qualify as intentional agents.

In the chapter I propose to focus on epistemic rationality, that is, the notion of rationality governing the relationship between what agents believe and the evidence available to them. In Sect. 2, by considering how widespread epistemically irrational beliefs are, I will suggest that

L. Bortolotti (✉)
Department of Philosophy, University of Birmingham, Birmingham, UK
e-mail: l.bortolotti@bham.ac.uk

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_12

epistemic rationality should not be seen as a condition for intentional agency, but rather as an aspiration for agents. I will point to examples suggesting that failures of epistemic rationality such as conservatism, superstition, and prejudice, are not a rare exception, but are frequent and systematic. Despite being epistemically irrational, behaviour can continue to be interpreted and predicted on the basis of the agents' intentional states.

In Sect. 3, I will suggest that, in some circumstances, some epistemically irrational beliefs are instrumental to people seeing themselves as intentional agents and increase their chances of fulfilling their goals as intentional agents. I will consider two cases: confabulated explanations of one's choices, and optimistically biased beliefs about the self. When people confabulate in explaining their choices, they tell more than they can know. In their self-related beliefs and predictions, they are more optimistic than is warranted by the evidence. In both cases, there is a failure of epistemic rationality. But the illusions people fall prey to allow them to see themselves as more competent and coherent than they actually are, and make it more likely that their motivation to act is sustained in the face of challenges and set-backs. The mere fact that they persevere in the pursuit of their goals increases their likelihood of fulfilling those goals, sometimes leading to success.

Epistemic irrationality can be both a curse and a blessing. By becoming aware of its pervasiveness and embracing it as an integral part of intentional agency in humans, we can start exploring both its dark and its bright side, limit its costs and find alternative ways to reap its benefits.

2 The Dark Side of Epistemic Irrationality

Typical humans are intentional agents, that is, their behaviour can be explained and predicted by attributing to them intentional states such as beliefs, desires, and emotions. Intentional agency can be assessed on the basis of standards of rationality. Calling an agent or a pattern of behaviour *rational* is a value judgement, and the reasons why the agent or the pattern of behaviour is valued may vary considerably across contexts (Bortolotti 2014).

This is because irrationality takes many forms. Agents are judged as irrational when their behaviour defies expectations and becomes hard to explain or predict; when their decisions are driven by emotions or instinct rather than by reflective deliberation; when their reasoning fails

to conform to basic principles of logic and probability; when their beliefs are inconsistent or badly supported by evidence; when their actions and thoughts are self-defeating and undermine their wellbeing; and the list could continue. Are there norms of rationality that cannot be violated without compromising agency itself? Some philosophers have argued that there are: if one's behaviour does not comply with those norms, then one cannot be ascribed intentional states or qualify as an intentional agent (Davidson 2004; Dennett 1971, 1989).

Which norms are the basic ones is open to debate, but I propose that we work here with epistemic rationality. Epistemic rationality requires that (1) beliefs are well-supported by the relevant evidence already at one's disposal, and (2) beliefs are responsive to relevant new evidence that becomes available. Evidential support and responsiveness to evidence are often conducive to the truth of one's beliefs, but they do not imply truth. It may seem plausible and it is certainly appealing to suppose that, in order to be ascribed intentional states and qualify as an intentional agent, the basic requirements of epistemic rationality must be met. One may ask what happens when a person violates the requirements of epistemic rationality. If her beliefs are not supported by evidence and are not responsive to evidence, can she coordinate effectively with other agents? Is she able to select the best means to achieve her goals? Can she plan effectively for her future? The answer, I believe, is yes. Both everyday experience and psychological evidence suggest that the requirements of epistemic rationality are not met by people's believing practices, and this has no particularly dire consequences for their intentional agency (Bortolotti 2010).

There are many strongly-held beliefs that seem to be based on no evidence or poor evidence, among which beliefs that we regard as superstitious (Vyse 2013). College students have superstitious beliefs about how to pass exams (they need to dress down), gamblers perform rituals in order to bring good luck to themselves (they think that dropping the dice is a bad omen), and people in show-business or in elite sport have their own superstitious beliefs and practices. One example of a belief that is widespread, has no evidence supporting it, and openly conflicts with other well-supported beliefs, is that more crimes, accidents, and psychological crises occur during nights of full moon (for recent discussions of this belief, see Hammond 2016; Lilienfeld and Arkowitz 2016). Among those endorsing this belief, we find police officers, medics working in emergency rooms, and mental health workers. This is surprising because

people in such categories have both practical experience and theoretical knowledge about the nature and the likely causes of emergencies in their respective fields. Even experts are not immune from the belief that a full moon negatively affects human behaviour.

The next two examples are about dismissing data against an accepted belief. If people already endorse a belief, then they are likely to be very critical of evidence that emerges against that belief, and they tend to dismiss the evidence. This happens when people are strongly motivated to accept the content of the belief, but also in circumstances in which there is no personal investment in the belief, apart from the mere fact that the belief has been adopted. When new data are consistent with the person's preferred theory, she tends to regard the data as robust and convincing. But when the new data are inconsistent with her preferred theory, she tends to regard the data as unconvincing. Typically, agents do not realise the extent to which their commitment to a preferred theory influences their assessment of the new data.

In a very interesting study, undergraduate students were asked to read reports of data relevant to whether the extinction of dinosaurs was due to volcanic eruptions (Chinn and Brewer 1993). Research participants read about the initial theory (e.g. "Dinosaurs went extinct due to volcanic eruptions"), which was well-argued for and contained many relevant pieces of evidence. Then they were asked to rate their belief in the theory, and most of them became very confident that the theory was correct. Next, they were divided into two groups. In group one, research participants read new evidence contradicting the initial theory (e.g. "Dinosaurs did not go extinct due to volcanic explosions, because eruptions were frequent but gentle"), and then provided ratings and reasons for their ratings. In group two, they read about new data supporting the initial theory (e.g. "Dinosaurs went extinct due to volcanic explosions, because eruptions were frequent and violent"), and also provided ratings and reasons for ratings. Finally, both groups were asked to what extent the new data were inconsistent with the initial theory, and they provided both ratings and reasons for ratings.

The assessment of the data was significantly influenced by the initial theory as predicted on the basis of previous studies, but participants did not notice it. When the data were consistent with the initial theory, participants found the data convincing. When the data were inconsistent with the initial theory, they found the data unconvincing. But the reasons for the assessment of the data were not transparent to them and

were not reflected in the reasons they provided for their ratings. In this series of experiments, participants showed what we might want to call 'conservatism' with respect to their beliefs, and a number of other epistemic weaknesses: they almost never suggested counterexamples for theoretical claims based on data, but preferred to propose alternative causal explanations even if these were clearly underdeveloped due to lack of elaboration; they used underspecified reasons to deny the impact of data on theories, and dismissed evidence that conflicted with their preferred theory even if they did not have good grounds to do so.

Another case of dismissing data against an accepted belief comes from the phenomenon of tableside racism, and more generally from cases of prejudiced belief based on negative stereotypes. Influenced by existing prejudice in the workplace, white waiting staff in US restaurants tend to form tenacious beliefs about the behaviour of black patrons. A common belief is that black patrons do not tip at all or do not tip generously. When such beliefs are disconfirmed, and black patrons do tip well, then the 'surprising' event is attributed to the waiter's excellent performance and does not convert into a piece of counter-evidence that shakes the initial belief (Fiske 2000).

Rusche and Brewster studied the behaviour of a small sample of white servers with respect to black patrons in large US restaurants, and found a variety of firmly held and widespread beliefs which affected the service. By means of field work in 2002 and in-depth questionnaires in 2004, the study revealed that the great majority of white servers in the sample believed that black patrons have unreasonable expectations, treat servers badly, and do not tip. For these reasons, white servers often avoid attending black patrons' tables, and share racist comments with other servers, often by using a code (Rusche and Brewster 2008). This form of tableside racism involves beliefs that are formed on the basis of very limited evidence or no evidence at all, as that new servers are 'brought up' into the racist culture by their more experienced colleagues and often form racist beliefs about black patrons in the absence of personal evidence. Moreover, the presence of obstinate beliefs helps perpetrate prejudice because the prediction that black patrons will not tip generously turns into a self-fulfilling prophecy.

> We argue that white customers are expected to leave fair tips and are therefore given service that merits them. Black customers, on the other hand, are expected to tip poorly and are therefore given poor service that merits

> bad tips. These performance expectations can thus become self-fulfilling prophecies if expectations shape service quality, and service quality influences tip size. In other words, if servers anticipate a poor tip, they may receive a poor tip, not because the customers are inherently bad tippers, but because they were given inferior service. In short, server discrimination is, in part, a function of the interaction between servers' cognition and the social climate in which they work. (Rusche and Brewster 2008, p. 2021)

The behaviour of people who hang onto beliefs that have no empirical support and that conflict with other things they know, uphold a theory that is disconfirmed by experts or by the data, and refuse to change their beliefs in the face of powerful counter-evidence is epistemically irrational. But, when people exhibit the epistemic irrationality I described, they do not stop behaving in a way that can be made sense of by attributing intentional states to them. They are still agents, and intentional agents whose behaviour can be explained and rationalised on the basis of their beliefs ("The gambler is expecting to lose because she dropped the dice"; "The waiter did not want to serve those black patrons because he thought that they would not tip him well"). Moreover, as we shall see in the next section, despite the epistemic irrationality of some of their beliefs, people can still be successful as agents, attaining goals that they find desirable.

3 The Bright Side of Epistemic Irrationality

Inspired by Eva's work on the role of rationality in the philosophy of Davidson (e.g., Picardi 1982, 1989, 1992), I dedicated my doctoral research and several years of my post-doctoral research to developing arguments for the peaceful co-existence of irrationality and intentional agency (e.g. Bortolotti 2004, 2005a, b). My most recent work, though, has turned to another aspect of the relationship between irrationality and human agency. After suggesting that not even systematic violations of basic standards of epistemic rationality rule out intentional agency, I became interested in whether some forms of epistemic irrationality support intentional agency and contribute to agents fulfilling their goals. I will refer to two cases of epistemically irrational behaviour that are very common, but until recently did not feature prominently in the philosophical literature on rationality and agency: confabulating explanations of one's choices; and having unrealistically optimistic beliefs about oneself or one's future.

Confabulated Explanations

When people confabulate they tell a story that they believe to be true, with no intention to deceive, but the story they tell is not based on evidence. One common situation in which people confabulate is when they explain their choices on the basis of facts that are plausible reasons for the choices they made, but that did not actually play a causal role in bringing about those choices. Confabulation can be observed in a variety of contexts, including consumer choice (Nisbett and Wilson 1977), moral (Haidt 2001) and political preferences (Hall et al. 2012), hiring decisions (Sullivan-Bissett 2015), and attitudes concerning romantic relationships (Wilson and Kraft 1993).

In their very influential 1977 study, Nisbett and Wilson argued that people are not aware of their mental processes when they give reasons for their choices. In a mall, participants were asked to choose some items as part of a consumer survey. Some participants were asked to choose between four different nightgowns. Other participants were asked to choose between four identical pairs of nylon stockings. Then, they were all asked why they made their choices. Experimenters found that participants' choices were very heavily influenced by the position of the items, and the item that was most on their right was the one most systematically preferred. But when people offered reasons for their choices, they mentioned features of the items they preferred, such as softness or colour, even when the items they chose from differed only in their position.

According to Nisbett and Wilson, people have no access to some of their mental processes, and thus they cannot always accurately explain why they made a certain choice. However, when they are asked for an explanation, they confabulate one. The reasons they provide for their choices are plausible given their background beliefs about what makes the chosen items preferable, but happen not to be part of the causal story leading up to their choices. Philosophers disagree about what is wrong with confabulation (see Lawlor 2003; Scaife 2014; Sandis 2015; Strijbos and Bruin 2015), but I suggest that the phenomenon that Nisbett and Wilson studied has three main features, also shared by other cases of everyday or mundane confabulation: ignorance of some of the key causal factors leading to the making of the choices (e.g. people not realising that they are affected by position effects in consumer choice); the production of an ill-grounded but often plausible explanation for the choices (e.g. people saying "I chose this pair of stockings because it is

brighter than the alternatives"); and often, but not always, the commitment to a further belief about the self or the world that does not fit the evidence in the specific case (e.g. people believing that the chosen pair of stocking is the brightest when it is identical to the alternatives).

Given the three features above, confabulation implies a failure of epistemic rationality. It may not be negligent to ignore some of the factors responsible for one's choices if these are opaque to introspection and can only be disclosed by acquaintance with experimental psychology. However, when people confabulate they *tell more than they can know*, and thus offer ill-grounded causal claims as explanations for their choices. They then can also end up committing to beliefs that are false and that are disconfirmed by evidence available to them.

Would it be possible to acknowledge ignorance instead of producing an ill-grounded explanation? In some circumstances, when a choice is driven by priming effects, the causal history behind the choice may not be easily accessed by introspection or inferred from behaviour. Moreover, confabulation experts agree that, when people confabulate explanations for their choices, they do not realise that they do not know the mental processes responsible for their choices (see Hirstein 2006; Turner and Coltheart 2010) and genuinely believe that the explanations they provide are accurate. That is one important aspect that distinguishes confabulation from deception. If it is true that when people confabulate they do not know that they do not know, then in genuine instances of confabulation people are in no position to acknowledge ignorance.

But suppose people could come to know about the priming effects driving their action, for instance, by studying experimental psychology. Would they be better off if, instead of confabulating, they offered an accurate explanation for their choices? Despite the epistemic costs I listed earlier, confabulating an explanation can have some psychological advantages over acknowledging ignorance or providing a better-grounded explanation and those advantages impact positively on agency. I am going to briefly consider three effects of confabulation.

First, confabulation can be *an act of self-enhancement*. Compared to replying "I don't know", which, as an answer to a self-related question, may undermine confidence and incur social sanctioning, confabulating an explanation can support one's sense of oneself as a competent decision maker (e.g. "Of course I can tell why I chose this pair of stockings!"). This is especially relevant when the choice is about a moral or

political preference, one's future career, or one's romantic partner: it is important to be able to explain a preference that can be self-defining.

Second, confabulation helps impose *coherence*. By confabulating, people make sense of behaviour that would otherwise lack an explanation and identify connections between their choices that make their various commitments more meaningful to them. The single choice becomes part of a largely coherent pattern of preferences that contributes to their image of themselves and can guide future behavior (e.g. "I am the sort of person who loves bright colours."). Again, this is more significant when people express a preference for a moral view or a political party than when they choose between pairs of stockings, because the former choices can be seen as making a more substantial contribution to their sense of self.

Third, a further advantage of confabulating over replying "I don't know" consists in allowing the *conversation* about the choice to continue, encouraging self-reflection and external feedback on either the choice or the reasons for it. When confabulated explanations of the plausible sort are shared, they are described as beneficial for social functioning.

It is worth noting that confabulating an explanation, with its positive psychological effects, may not only be advantageous with respect to acknowledging ignorance, but also with respect to offering a better-grounded explanation, at least in the short term (Bortolotti 2018a). Explaining choices as the effect of priming effects implies a recognition that such choices were not entirely under the person's conscious control and may either weaken the person's sense of self as a competent and coherent agent or fail to encourage self-reflection or external feedback on the choice.

Optimistically Biased Beliefs

Positive illusions are systematic tendencies, widespread in the non-clinical population, to form beliefs and make predictions that are overly optimistic. In the classic literature three types of biases are described: the *illusion of control*, that is, the fact that people believe that they can control events that are independent of them; the *superiority illusion*, that is, the fact that people believe that they are better than average at most things; and the *optimism bias*, that is, the fact that people expect their future to be rosier than warranted by the evidence or rosier than the future of other people.

There are numerous examples of everyday behaviours that are guided by such biases. For instance, in a casino people tend to think that they have a better chance at winning when they are the ones rolling the dice, and they bet more money in those circumstances (Vyse 2013) because they have an illusion of control. When college professors in the US were asked whether they did above-average work, 94% of them say they did (Cross 1977) which suggests that at least some of them overestimated the value of their work and were vulnerable to the superiority illusion. Finally, people tend to underestimate the likelihood that their marriage will end in divorce or that they will develop a serious health condition during their lives (Sharot et al. 2011), and when new evidence relevant to such circumstances becomes available they tend to update their beliefs accordingly if the new evidence supports an optimistic outcome, but they are reluctant to update their beliefs otherwise.

Unrealistically optimistic beliefs and predictions are a perfect example of epistemically irrational cognitions, as they can both lack evidential support and be resistant to counter-evidence. They are likely to be due to cognitive and motivational factors, including incompetence, neglect of relevant information, failure to learn from feedback, selective updating, and defensiveness (Alicke and Sedikides 2011). But, despite their epistemic irrationality, optimistically biased beliefs seem to have significant benefits in several areas, including romantic relationships and health. Let us see how optimism affects relationships first. It was found that even those who are well informed about divorce rates in the society where they live tend to underestimate their likelihood to get separated or divorced, and overestimate the longevity of their relationships (Baker and Emery 1993). They are subject to the *love-is-blind illusion* in that they tend to be blind to the faults of their romantic partners, and perceive their partners as better than average in a number of domains including intelligence and attractiveness (Murray and Holmes 1997). People also view their actual partners as more similar to their ideal partners than the partners are (Murray et al. 1996). Finally, people tend to perceive their relationship as better than most relationships (Buunk and Eijnden 1997).

Such biases are correlated to better responses to stressful situations and to conflict in long-term relationships: unrealistically optimistic couples seem to cope better (that is, in more constructive ways) with adversities. This effective coping is correlated with greater relationship satisfaction and stability over time. Moreover, in an influential study (Murray et al. 1996) it was suggested that “intimates who idealise one

another appear more prescient than blind, actually creating the relationships they wish for as romances progress" (p. 1155), the idea being that people who are being idealised by their partners make a genuine attempt at living up to the high standards.

The study of the effects of optimism on health prospects tells a similar story. People have optimistically biased beliefs concerning their health, and such beliefs often motivate them to engage in health-promoting behaviour. One example from the optimism bias literature that is quite striking concerns seropositive men who were found to be significantly more optimistic about not getting AIDS than seronegative men (Taylor et al. 1992). Their belief that they would not get AIDS seemed to be a consequence of their discovery of their seropositive status, and part of their coping mechanisms.

In a paper discussing the benefits of positive illusions (Taylor and Brown 1994), the optimistic predictions of the seropositive men were "associated with reduced fatalistic vulnerability regarding AIDS, with the use of positive attitudes as a coping technique, with the use of personal growth/helping others as a coping technique, with less use of avoidant coping strategies, and with greater practice of health-promoting behaviors" (p. 24). Another aspect of the relationship between optimism and health promotion concerns the often illusory belief that people can control their health. When women diagnosed with breast cancer adopt self-enhancing beliefs, such as "I am stronger as a result of the illness", "I can cope better than other cancer patients", "I can control my health condition from now on" (Taylor 1983), they cope better with their situation and behave more responsibly because they believe that they have the power to avoid or delay illness (Taylor and Sherman 2008).

Indeed, people who are optimistic about their health prospects are found to make greater life-style changes in order to maintain their health. Two common explanations for this outcome point to stress management and coping: optimistic people manage stress more effectively, and this in turn allows them to sustain the motivation to engage in health-promoting behaviour; and they react to a critical health condition by facing rather than by avoiding problems.

In the two cases above, seropositive men believing that they can avoid AIDS and women with a history of breast cancer believing that they can avoid the return of the illness, optimism leads people to make life-style changes that help them restore or maintain good health. Such evidence concerning optimism and health promotion, combined with

the evidence of the correlation between optimism and lasting, satisfactory relationships, suggests that epistemically irrational beliefs have a bright side. In particular contexts, having an overly optimistic assessment of the situation at hand helps engage with the situation in a more productive and resilient way, enabling people to attain some desirable goals (Bortolotti 2018b).

4 Conclusions

Epistemic rationality is often considered a condition for intentional agency but evidence disclosing systematic tendencies to adopt epistemically irrational beliefs that are often acted upon suggests that standards of epistemic rationality can be and often are violated without compromising agency. We reviewed some of the relevant evidence in Sect. 2. Philosophers are so attached to the claim that humans are rational that they have often attempted to minimise the impact of the psychological evidence on claims about human capacities. They have argued that even locally irrational agents are rational *by and large* or that, when the rationality constraint seems to fail, it is because we are not working with the right notion of rationality. Both moves are unconvincing in the light of how pervasive and epistemically problematic conservatism, superstition, and prejudice really are. Even on those accounts where the only form of rationality that counts as a condition for intentional agency is coherence in the belief system, the phenomena we described would raise concerns. Superstition and prejudice can be characterised as a clash between a person's commitments: the doctor who expects more accidents in a night of full moon is likely to have conflicting beliefs about what causes accidents; the waiter who explains the generous tips received from white and black patrons differently is likely to take very similar situations as evidence for different claims for no good reason.

Superstitious and prejudiced beliefs are harmful in epistemic and non-epistemic ways, and should be avoided. Restoring epistemic rationality in the contexts of the lunar effect and tableside racism is an important undertaking, as it is in other contexts where superstition and prejudice are widespread. But in Sect. 3 we saw how some epistemically irrational beliefs have psychological benefits which impact on agency and success. Confabulated explanations of one's choices are always ill-grounded and often inaccurate, but support the perceived agency and motivation of the people who confabulate, by promoting

self-enhancement, facilitating psychological integration, and enabling feedback from peers. Confabulators impose coherence on their choices and benefit from an illusion of competence that increases their sense of agency. Optimistically biased beliefs about oneself or one's future are not based on evidence and often turn out to be false, but are correlated with people's success in some domains, in that having the beliefs makes people more likely to attain desirable goals. Optimists show greater resilience and cope more constructively with adversities than realists and this makes them more likely to have long-lasting and satisfying romantic relationships and to enjoy good health. The key concept in the counter-intuitive idea that some epistemic irrationality breeds success is *motivation*. When choices are integrated in a pattern of behaviour that makes sense and can be shared, people see themselves as competent and coherent. Thanks to overly optimistic beliefs and predictions, agents become more productive, more resilient, better at planning, and more effective at problem-solving (Alicke and Sedikides 2011). In both cases the illusion of agency make people behave more like agents.

The epistemically irrational beliefs we described have many shared characteristics. They start out as lacking evidential support and they are often strenuously resistant to counter-evidence. In addition, they may avoid disconfirmation by giving rise to self-fulfilling prophecies: the black patron who experiences bad service may not tip generously and the romantic partner who is idealised may live up to the ideal standards. Epistemically irrational beliefs, whether psychologically beneficial or not, compromise our understanding of the world around us, at times in unhelpful ways causing confusion and division, and at other times in helpful ways strengthening our belief in our own efficacy. The challenge is to reduce the evident costs of epistemic irrationality for agency without canceling its potential benefits, and this can be done by accepting a psychologically realistic notion of what human agents are like, and interrogate evidence about the relationship between epistemic and psychological benefits whenever possible.

Note

1. The author acknowledges the support of the European Research Council under the Consolidator grant agreement number 616358 for a project called Pragmatic and Epistemic Role of Factually Erroneous Cognitions and Thoughts (PERFECT).

References

Alicke, Mark D., and Constantine Sedikides. 2011. *Handbook of Self-Enhancement and Self-Protection*. New York: Guilford Press.

Baker, Lynn A., and Robert E. Emery. 1993. When Every Relationship Is Above Average. *Law and Human Behavior* 17 (4): 439–450. https://doi.org/10.1007/BF01044377.

Bortolotti, Lisa. 2004. Can We Interpret Irrational Behavior? *Behavior and Philosophy* 32 (2): 359–375.

Bortolotti, Lisa. 2005a. Delusions and the Background of Rationality. *Mind & Language* 20 (2): 189–208. https://doi.org/10.1111/j.0268-1064.2005.00282.x.

Bortolotti, Lisa. 2005b. Intentionality Without Rationality. *Proceedings of the Aristotelian Society (Hardback)* 105 (1): 369–376. https://doi.org/10.1111/j.0066-7373.2004.00119.x.

Bortolotti, Lisa. 2010. *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Bortolotti, Lisa. 2014. *Irrationality*, 1st ed. Malden, MA: Polity Press.

Bortolotti, Lisa. 2018a. Stranger than Fiction: Costs and Benefits of Everyday Confabulation. *Review of Philosophy and Psychology* 9 (2): 227–249.

Bortolotti, Lisa. 2018b. Optimism, Agency, and Success. *Ethical Theory and Moral Practice*.

Buunk, Bram P., and Regina J.J.M. van der Eijnden. 1997. Perceived Prevalence, Perceived Superiority, and Relationship Satisfaction: Most Relationships Are Good, But Ours Is the Best. *Personality and Social Psychology Bulletin* 23 (3): 219–228. https://doi.org/10.1177/0146167297233001.

Chinn, C.A., and W.F. Brewer. 1993. The Role of Anomalous Data in Knowledge Acquisition: A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research* 63 (1): 1–49. https://doi.org/10.3102/00346543063001001.

Cross, K. Patricia. 1977. Not Can, But Will College Teaching Be Improved? *New Directions for Higher Education* 1977 (17): 1–15. https://doi.org/10.1002/he.36919771703.

Davidson, Donald. 2004. *Problems of Rationality*. Oxford: Clarendon Press.

Dennett, D.C. 1971. Intentional Systems. *The Journal of Philosophy* 68 (4): 87–106. https://doi.org/10.2307/2025382.

Dennett, Daniel Clement. 1989. *The Intentional Stance*. Cambridge, MA: MIT Press.

Fiske, Susan T. 2000. Stereotyping, Prejudice, and Discrimination at the Seam Between the Centuries: Evolution, Culture, Mind, and Brain. *European Journal of Social Psychology* 30: 299–322.

Haidt, J. 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108 (4): 814–834.

Hall, Lars, Petter Johansson, and Thomas Strandberg. 2012. Lifting the Veil of Morality: Choice Blindness and Attitude Reversals on a Self-Transforming Survey. *PLoS ONE* 7 (9): e45457. https://doi.org/10.1371/journal.pone.0045457.

Hammond, Claudia. 2016. Does a Full Moon Make People Mad? http://www.bbc.com/future/story/20131029-does-a-full-moon-make-people-mad. Accessed 7 Nov.

Hirstein, William. 2006. *Brain Fiction: Self-Deception and the Riddle of Confabulation: Self-Deception and the Riddle of the Confabulation*, New ed. Cambridge, MA: MIT Press.

Lawlor, Krista. 2003. Elusive Reasons: A Problem for First-Person Authority. *Philosophical Psychology* 16 (4): 549–564. https://doi.org/10.1080/0951508032000166969.

Lilienfeld, Hal, and Scott O. Arkowitz. 2016. Lunacy and the Full Moon. *Scientific American*. https://www.scientificamerican.com/article/lunacy-and-the-full-moon/. Accessed 7 Nov.

Murray, Sandra L., and John G. Holmes. 1997. A Leap of Faith? Positive Illusions in Romantic Relationships. *Personality and Social Psychology Bulletin* 23 (6): 586–604. https://doi.org/10.1177/0146167297236003.

Murray, Sandra L., John G. Holmes, and Dale W. Griffin. 1996. The Self-Fulfilling Nature of Positive Illusions in Romantic Relationships: Love Is Not Blind, But Prescient. *Journal of Personality and Social Psychology* 71 (6): 1155–1180. https://doi.org/10.1037/0022-3514.71.6.1155.

Nisbett, Richard E., and Timothy D. Wilson. 1977. Telling More than We Can Know: Verbal Reports on Mental Processes. *Psychological Review* 84 (3): 231–259. https://doi.org/10.1037/0033-295X.84.3.231.

Picardi, Eva. 1982. Interpretazione Radicale e Teorie Della Verità. *Lingua e Stile* XVIII (1): 51–67.

Picardi, Eva. 1989. Davidson on Assertion, Convention and Belief. In *The Mind of Donald Davidson*, ed J. Brandl and W.L. Gombocz. Amsterdam: Rodopi.

Picardi, Eva. 1992. Credenze E Razionalità. *Iride* 10 (1992–1993): 177–181. *Iride* 10: 177–181.

Rusche, Sarah Nell, and Zachary W. Brewster. 2008. 'Because They Tip for Shit!': The Social Psychology of Everyday Racism in Restaurants. *Sociology Compass* 2 (6): 2008–2029.

Sandis, Constantine. 2015. Verbal Reports and 'Real' Reasons: Confabulation and Conflation. *Ethical Theory and Moral Practice* 18 (2): 267–280. https://doi.org/10.1007/s10677-015-9576-6.

Scaife, Robin. 2014. A Problem for Self-Knowledge: The Implications of Taking Confabulation Seriously. *Acta Analytica* 29 (4): 469–485. https://doi.org/10.1007/s12136-014-0226-8.

Sharot, Tali, Christoph W. Korn, and Raymond J. Dolan. 2011. How Unrealistic Optimism Is Maintained in the Face of Reality. *Nature Neuroscience* 14 (11): 1475–1479. https://doi.org/10.1038/nn.2949.

Strijbos, Derek, and Leon de Bruin. 2015. Self-Interpretation as First-Person Mindshaping: Implications for Confabulation Research. *Ethical Theory and Moral Practice* 18 (2): 297–307. https://doi.org/10.1007/s10677-015-9579-3.

Sullivan-Bissett, Ema. 2015. Implicit Bias, Confabulation, and Epistemic Innocence. *Consciousness and Cognition* 33 (May): 548–560. https://doi.org/10.1016/j.concog.2014.10.006.

Taylor, S.E. 1983. Adjustment to Threatening Events: A Theory of Cognitive Adaptation. *American Psychologist* 38: 1161–1173.

Taylor, Shelley E., and Jonathon D. Brown. 1994. Positive Illusions and Well-Being Revisited: Separating Fact from Fiction. *Psychological Bulletin* 116 (1): 21–27.

Taylor, S.E., M.E. Kemeny, L.G. Aspinwall, S.G. Schneider, R. Rodriguez, and M. Herbert. 1992. Optimism, Coping, Psychological Distress, and High-Risk Sexual Behavior among Men at Risk for Acquired Immunodeficiency Syndrome (AIDS). *Journal of Personality and Social Psychology* 63 (3): 460–473.

Taylor, S.E., and David K. Sherman. 2008. Self-Enhancement and Self-Affirmation: The Consequences of Positive Self-Thoughts for Motivation and Health. In *Handbook of Motivational Science*, ed. James Y. Shah and W.L. Gardner, 57–70. New York: Guilford Press.

Turner, Martha, and Max Coltheart. 2010. Confabulation and Delusion: A Common Monitoring Framework. *Cognitive Neuropsychiatry* 15 (1): 346–376. https://doi.org/10.1080/13546800903441902.

Vyse, Stuart A. 2013. *Believing in Magic: The Psychology of Superstition*, Updated ed. Oxford: Oxford University Press.

Wilson, Timothy D., and Dolores Kraft. 1993. Why Do I Love Thee?: Effects of Repeated Introspections About a Dating Relationship on Attitudes Toward the Relationship. *Personality and Social Psychology Bulletin* 19 (4): 409–418. https://doi.org/10.1177/0146167293194006.

Reasons and Causes in Psychiatry: Ideas from Donald Davidson's Work

Elisabetta Lalumera

1 Introduction

I am cycling, I face a turn, I want to turn right and I know how to signal it—that is, by extending my right arm; so I extend my right arm. That I want to turn right, to signal it, and I know how to do it is the reason why I move my arm as I do. A car driver behind me, assuming that she is familiar with the basic conventions of cycling on a public road, can understand my behaviour and evaluate my action as correct. But what caused my extending of my right arm? Or, in other words, what is the best causal explanation of my movement? We are inclined to respond by describing a physiological mechanism in my body involving brain systems (cognition, volition, motor-control), and the muscular system. Still, I stretched out my right arm for a reason, that is, that I wanted to signal my right turn. What is the relationship between the physiological mechanism, and my reason?

Consider now a more complex scenario. I am running for mayor in my town, and I am very close to being elected. In the final phase of the

E. Lalumera (✉)
University of Milano-Bicocca, Milan, Italy
e-mail: elisabetta.lalumera@unimib.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_13

campaign, I am accused by one of my political rivals of being involved in an illicit business, and the police start investigating. My name appears in the local news for days, my political career is over, and I am publicly humiliated. I start to feel deeply sad and hopeless, and to lose interest in my life activities and relationships. After some months, I am proved innocent, but my state of depression persists. During psychotherapy, it comes out quickly that the humiliation I felt, considered within my own set of expectations and values, is the reason why I am depressed—the reason that triggered my condition. Nevertheless, as depression is a mental disorder, the explanation of my depression involves an etiological account of brain mechanisms, the dysfunction of which causes my state. What is the relationship between the brain mechanisms and my reason? How can the latter *trigger* a change in the former?

The relationship between reasons and causes is key to philosophical as well as scientific discussions on the status of psychiatry and psychopathology in general (Thornton 2007). On one side of the divide, supporters of a strong version of the medical model applied to mental disorders call for explanations in terms of neuroscience, biology, genetics, or a combination of these levels, claiming that psychiatry should search for causes by going for the etiology of mental disorders (Insel and Cuthbert 2015; Murphy 2006). On the other side, some psychoanalysts, phenomenologists, psychotherapists and clinical psychologists of various orientations approach the explanation and treatment of mental disorders via the interpretation of patients' array of reasons, broadly conceived, including desires, beliefs, convictions, affective states, norms and values (see e.g. Stanghellini and Fuchs 2013).

The divide between causal approaches and reason-based approaches in psychiatry (or between *explanation* and *understanding* of mental disorders) has cultural and historical roots that trace back to the origins of psychiatry and psychology as sciences (Jaspers 1923/1963), and they deepened during the cultural battle of the '70s and '80s of the last century. In those years, from one side, psychiatry and experimental clinical psychology were accused of medicalizing the problems of life while, from the other side, psychotherapies were stigmatized as scientifically and methodologically inadequate.[1]

Though it is implausible that such strong roots of antagonism between causal and reason-based approaches could be eradicated by philosophical arguments alone, there are other forces at play, such as the shift in scientific research paradigms, which may contribute to bridging the

divide. Some of the contemporary approaches to scientific psychiatry now de facto consist of multi-factorial or mixed models of the explanation of mental disorders, where *reasons*[2] such as unfortunate life events (for example, a public humiliation, difficulties in family and affective relationships) enter as *causes* of disorders (such as depression, substance abuse, and eating disorders), along with brain dysfunctions and genes (Kendler et al. 1999; Tennant 2002). In order to make room for these new models, there is the need of a conceptual clarification of the notions involved, namely reason, cause and the causation relation. Which notions of cause, reason and causation, if any, can fit into multi-factorial explanatory models of psychiatric disorders?

This paper attempts to answer that question. There are basically two parts of the same agenda that have to be completed. The first is to reflect on the notion of causality, so as to accommodate heterogeneous factors. John Campbell and Kenneth Kendler recently defended an interventionist account of causation, where to cause is to make a systematic difference on a given variable, rather than 'to bring it about' physically; as interventionism allows for heterogeneous factors such as life events and socioeconomic status to feature as causes of mental conditions, it is appropriate for multi-factorial explanations of psychiatric disorders (Campbell 2007; Kendler and Campbell 2009).[3]

The other part (not an alternative one, but a complementary one) is to discuss the ontology of reason and causes, and defend the claim that reasons can be causes. This paper takes this second route. Some authors in the recent debate take the claim for granted without argument—notably, for example, Dominic Murphy writes that 'most philosophers now are happy to think of reasons as causes' (Bolton 1997a; Murphy 2013, p. 983).[4] But the *locus classicus* of the claim that reasons are causes, however, is the work of Donald Davidson, situated within the philosophical debate on intentional action, which took place roughly from the 1950s to the 1970s[5] (Davidson 1963; Davidson 2001). When Davidson wrote, the consensus was that rationalizing explanations cannot be causal explanations, logical and normative relations among reasons being different and irreducible to causal relations. Davidson's seminal paper 'Actions, Reasons, and Causes' and the discussion that followed significantly contributed to bridging the divide—and actually reversed the consensus, in the philosophy of action, from rationalism to causalism.[6]

I think that Davidson's articulation of the idea that reasons are causes—with the necessary adaptations to a different set of questions, and with a broadened notion of reason, which will be specified in Sect. 3 below—can be fruitfully applied to the contemporary debate in the philosophy of psychiatry and be part of projects of clarification of multifactorial models of the explanation of mental disorders.

This is how this paper is organized. In Sect. 2, I present some facts and pragmatic considerations that may count as advantages of multifactorial explanatory models of mental disorders—where reasons are causes—when compared to strong medical models (to use Dominic Murphy's well-known expression) and to purely hermeneutic ones. In Sect. 3, I present what I take to be the Davidsonian insights, which are relevant to the debate on reasons and causes in psychiatry; I also respond to a worry about extending the meaning of 'reason', and to an objection raised to Davidson by Tim Thornton (2007). Section 4 contains my concluding remarks.

2 Why We Need an Account of Reasons as Causes in Psychiatry

Why should we want reason-based explanations to be also causal explanations—why should we want reasons to be causes at all? What is wrong with the divide? In this section I present two answers, pertaining to the problem of the explanation of mental disorders.

The first answer is a fact, rather than an argument. From the point of view of psychiatric research, only a few mental disorders[7] can be adequately explained by one or a limited set of causal factors, such as a brain lesion or an infection from a pathogenic agent: head injuries or strokes can lead to psychological dysfunctions, and bacterium *Treponema pallidum* is the cause of syphilis, which may have symptoms of dementia.[8] Apart from these few examples, empirical studies show that many mental disorders are better explained by various heterogeneous factors, some of which are at the micro level (brain chemistry, genetics), but others of which are at the macro level (psychological, social, economical and cultural). This suggests the adoption of heterogeneous or pluralistic explanatory models, rather than reductionist or strictly internalist (brain-based) models, as philosophers of science have highlighted (see e.g. Campaner 2014; Sullivan 2016).

Here are two specific examples. Research on alcohol disorders shows that all the following kinds of factors are significant: biological and genetic (genetic effect on alcohol metabolism and with systems that interact with ethanol, genetic liabilities to psychoactive stimulations and substance dependence); psychological (neuroticism, impulsivity and alcohol expectancies); social (peer substance use, social class); and cultural (acceptability of alcohol abuse, gender stereotypes and social stereotypes) (Boden et al. 2014; Kendler 2008, p. 696). The second example is depression.[9] A longitudinal studies on female twins, with data collected over 10 years, found that all the following conditions were significantly correlated with the insurgence of the disorder: genetic risk, disturbed family, childhood parental loss, neuroticism, low self-esteem, low education, low social support, divorce, marital problems and stressful life events (Kendler et al. 2002). Another study on male twins reported humiliation as a significant factor, as in the vignette I opened with in the Introduction (Kendler and Halberstadt 2013; for a review, see Tennant 2002).

When an empirical study finds that divorce or humiliation is significant for depression, it means that factor X is causally relevant to some occurrence Y. The general pattern is that X *makes a certain kind of difference* to the occurrence of Y in some reference class C. *Prima facie*, this is just a causal explanation, and it is how a causalist (in Jaspers's terms, an explanatory) account of mental disorders works. How are reasons supposed to fit in here? The point is that some of the factors replacing X—such as the early death of a parent, or a divorce, or a public humiliation—earn their place in studies on mental disorders only qua 'denizens of the realm of reasons'[10]: they can be relevant to depression or alcoholism only as meaningful contents, represented in people's minds as such, and that relate to their normative systems.

The point can also be put this way: by other descriptions, the death of a parent is just a date on the calendar, or a process involving a specific human body, and a divorce is a legal act, with no intelligible connection with another person's condition of depression. Why hypothesize that that kind of event or condition, and not any other (say, the possession of a driving licence, or the colour of one's eyes) makes a difference to depression, or alcoholism? Nothing else besides the acknowledgment that parental loss and divorce, and not, say, facts like the possession of a driving licence or having brown eyes, are connected in special ways with someone's systems of values and norms: they are meaningful. In other words, it is true that reason-based accounts tell us how it is rational and

appropriate to behave, and causal explanations specify which behaviour we should expect. However, in these cases the expected behaviour is such, because of the meaning and value of the events involved.

The conceptual point does not change if we shift focus from the individual case and generalize the result by examining populations or subpopulations of individuals, as in experimental studies. What empirical studies with twins track and measure is the effect of meaningful events qua meaningful in the life of subjects, not just of events under any description.

If these considerations are sound, when we hypothesize that an early parental loss or a divorce caused someone's depression, we are relating *reasons* and an organic condition in one and the same explanation. Thus, if we take explanatory pluralism in psychiatry seriously, we need an account of how reasons can be causes.[11] The question ought to shift from *whether* reason can be causes, to *in virtue of what* they can be.

A second consideration in favour of working with philosophical models that bridge the cause-reason divide has to do with the status of psychoanalysis, and more generally with the status of interpersonal talk-therapies. These approaches to the treatment of mental disorders work with reasons in the broad sense (internalized meaningful events, traumas, repressed thoughts, and desires). On the one hand, there is an influential tradition—the hermeneutic tradition—supporting the view that, when a psychotherapist individuates certain reasons as the origin of her patient's distressful state or problematic behaviour, she is neither discovering a cause nor explaining the nature of the patient's condition. Rather, the clinician proposes a narrative that interprets and understands the patient's feelings and doings and that the patient can in principle be able to recognize as fitting her case, and thus receive benefit from the insight (see e.g. Klein 1976; Ricoeur 1970). At least in part, the retreat of psychoanalysis (and psychodynamic approaches) from an inquiry into causes and explanations to an approach of narratives and interpretations was a reaction to the attacks from Grünbaum and other writers, who put into question (*inter alia*) the possibility of establishing causal claims between patients' systems of reasons and pathologies of behaviour, just on the basis of observation of clinical cases, with no proper experimental study with a comparative class. If psychoanalysis cannot be appropriately confirmed as a scientific theory about the functioning of human minds, then it would do best to be proud to be something radically different (Bolton 1997b; Grünbaum 2004).

On the other hand, however, at the level of population studies, recent meta-analyses on randomized-controlled trials show that interpersonal psychotherapies of various theoretical orientations are effective for some mental conditions, such as depression, anxiety, and eating disorders (Cuijpers et al. 2016; Fonagy 2015), and that they are more effective than placebo treatments (Cuijpers et al. 2014). That is to say, granted that the studies are appropriately conducted,[12] interpersonal psychotherapies make a difference on people's symptoms and prognosis. How can this be possible, even in principle, if the dynamics of reasons are not also a causal process? That reasons are also causes is the (likely, though defeasible) conclusion of an inference for the best explanation, from the premise that psychotherapies can be effective. This is not meant to suffice for a cheap vindication of the current status of psychoanalysis as a science, for qualms about confirmation remain in place about it, as they do about any theory of the dynamics of reasons in the human mind, for which the main evidence is clinical data (Fonagy 2003). However, the possibility of a multilevel scientific approach to mental disorders—where reason-explanation and therapy are employed along with causal explanation and therapy—should not be a priori undermined (Fonagy 2015; Lacewing 2013).

The thought that psychotherapies have a (often implicit) causal vocation can be rephrased as a point about the clinician-patient's relation. I quote from Derek Bolton:

> As a treating clinician, one presents one's practice as creating for the patient some possibilities of change, and this implicit communication to the patient (and to whoever is paying the bill) implies an interest in causal connections of the form: if I, or you, do such and such, the chance of change is increased. It may well be compatible with this to suppose that narratives created in psychotherapy are not themselves causal hypotheses, but there has to be even here, I suggest, a weaker claim that creating a coherent and credible narrative makes a difference to a person's level of distress, or to self-esteem, or to the feeling of self-control, or to something else that matters to the person. Again, some psychotherapists may say that they do not or cannot treat symptoms, but aim rather to make a difference to the way the person feels about having them, and this re-framing can reduce distress. But still this line of thought involves crucial assumptions to the effect that the meaningful makes a difference, and hence, apparently, is involved in causality. (Bolton 1997b, p. 274)

The bottom line of these considerations about talk-therapies is the same as the one I derived earlier from the examples of multi-factorial models of mental disorders: there is the need of a metaphysics of reasons and causes that abolishes the divide, for such a model is already implicit in the current psychiatric practice of research and treatment.

3 Davidson on Reasons as Causes

The conclusion of the previous section was that a philosophical articulation of the claim that reasons can be causes is needed in order to make sense of, and to clarify, two research projects in contemporary psychiatry: the development of multi-factor models of mental disorders and the integration of interpersonal psychotherapies within scientific practice and patients' care. In both cases, what we need is a philosophical legitimation of the claim that reasons can be causes. In this section, I argue that Davidson's ideas can be useful in this respect.

A preliminary objection needs to be addressed before starting. Davidson wrote in the context of the debate in the philosophy of action, where the notion of reason was meant to cover the content of *conscious* mental states, which can be used by the subjects in providing rationalizations of their own actions, or by some third part in making sense of them. What we need in the debate in the philosophy of psychiatry is a broadening of the basic notions involved, so that reasons comprehend unconscious or implicit contents, affective states, desires and normative commitments, and correspondingly behaviour comprises not just actions but also habits, moods and unconscious decisions, for these components have a place in the explanation of mental disorders. Are Davidsonian reasons, and the broadened notion of reason I have been employing, similar enough? I think so. The important divide, both in the philosophy of action and in the philosophy of psychiatry, is about the possibility of meaningful states to enter in causal relations in virtue of their meaning, though meaning confers normative relations, and normative relations are not identical to causal ones. In this sense, an explicit and conscious intention I may voice to myself while I cycle (to signal my right turn) is no different than the unconscious repressed desire that the therapist may individuate as the origin of a patient's condition: they are meaningful states, related with other meaningful states via normative or 'thematic'[13] relations.

With these considerations in place, here are Davidson's ideas that can be extrapolated from the debate in the philosophy of action and employed for defending a unitary account of the human mind. First, there is a point about explanation: of all the possible and good rationalizing explanations of a behaviour, only one is the reason why, namely, the cause of the behaviour. All the others can be good justifications or reconstructions, but they do not pick out what makes a difference for the person's behaviour. This is the (oversimplified version) of the main claim in 'Actions, Reasons, and Causes' (Davidson 1963).

Second, Davidson thought that a reason can be a cause in virtue of being one and the same thing, a physical thing—Davidson's favourite ontology included events—and he argued that a reason and a cause can be the same event, under different descriptions (Davidson 1970). The event is a physical particular: any physical particular would fit the pattern, if one were inclined towards a different ontology (this is the monist part of his 'anomalous monism', a token-identity theory).

Third, Davidson embraced a conceptual anti-reductionism of the mental to the physical: in semantic terms, we cannot systematically translate mental predicates into physical predicates. This is the anomalous part of his anomalous monism[14] (Yalowitz 2014). The contemporary version of this claim is that we cannot systematically relate psychological predicates into brain predicates, as in projects of reductive neuroscience—we cannot, for example, systematically relate 'being humiliated' with a pattern of brain activation (see Shea 2013, for a contemporary discussion within the debate on mental disorders).

Among the various important conceptual objections to Davidson's proposal, a key one for my concerns here is *epiphenomenalism* (Kim 2000). In its simplest form, the objection states that, if all causes are physical and all reasons are physical events, then reasons qua reasons are causally superfluous. Davidson's own reply involves, again, the notion of explanation: sometimes we need to describe an event or a physical particular via its mental (non-physical) properties, because explanations have *goals* and depend on our *interests,* and are adequate or not adequate in relation to them. In all cases where the goals and interests of our explanations are not set by the agenda of chemistry, physics and other basic sciences, a causal explanation would be inadequate if we replaced a mental predicate with a physical predicate (Davidson 1991).

This point is analogous to what I argued in Sect. 2 about multi-factor models of mental disorders. If we had a complete physical description of

an event such as the loss of a parent or a divorce (say, as a brain state in the memory system), that would not be an adequate psychiatric description of the cause of someone's distress or, more generally, mental condition for the goals of research and treatment. Discussing the appropriate kind of explanation for psychiatry, Kenneth Kendler describes the following scenario: a brilliant young man abandons a promising career in science because he wants to become a Catholic priest. The psychiatrist is consulted by the man's mother, who asks for some pharmacological remedy in order to intervene in her son's mental condition. Kendler argues that even though here we have 'a higher-order system that is completely constituted from lower-order elements'—the would-be priest's mind-brain—'an intervention at the level of the lower-order elements is likely to be, at best, inefficient and, at worst, ineffective and possibly harmful'. Here, reasons are causally relevant because they enter into the sole adequate explanation of what is happening to the person, and constitute the sole adequate level where intervention makes sense (Kendler 2005, p. 435).

Not all the details of Davidson's overall picture of the mind-brain problem, and of the philosophy of action, fit in the project of bridging the divide between reasons and causes in psychiatry. For example, as I noted in passing in the Introduction, some philosophers of psychiatry are inclined towards interventionist models of causation, whereas Davidson held that causation is always law-like (Davidson 1963, 1970). On the other hand, a feature that has traditionally puzzled philosophical commentators, the conceptual irreducibility of the mental to the physical, or the anomalous part of anomalous monism, is actually a strategic advantage in the specific concerns of psychiatric explanation.

Let me put the point another way. Even the most enthusiastic supporters of psychiatry as neuroscience acknowledge that we are now in a state of relative ignorance as to the systematic mapping of psychological predicates onto the brain—while some psychological constructs are relatively well translatable into brain predicates, there is no one-to-one correspondence, and some other, more complex predicates, such as 'being humiliated', are simply bad candidates for a reduction (Kendler 2005; Murphy 2006; Shea 2013). Surely, what for Davidson was an a priori impossibility now tends to be considered more as a huge practical difficulty. However, the point remains that we need to make sense of a causal impact of reasons in the causal world without the need of a translation (a reduction) of the psychological to the low-level physical, and an anomalous-monist framework makes that possible.

A final (and related) concern about the possibility of employing the main claims of the Davidsonian framework within the philosophy of psychiatry comes from Tim Thornton. In a critical paper of Derek Bolton's account of reasons and causes, he moves to Davidson the following objection:

> He provides no answer of why it is generally the case that the rational power and the causal power of reasons stand in proportion. There are simply no resources in his account to answer this question..[…] he fails to unite reason explanation and causal explanation. The fact that events can play a role in both spaces is not sufficient to unite reasons and causes. While his account manages to display the rational structuring of reason and give an account of the causal role of reasons, it cannot explain how reason can itself play a role in causal explanation. He fails to reconcile or unite the rational and the causal. (Thornton 2007, p. 310)

I think that Thornton is right: given the characteristics of anomalous monism, there is no a priori justification for the claim that 'it is generally the case that the rational power and the causal power of reasons stand in proportion'. However, this need not be an objection. It may be that some other theory has the resources to explain this fact, a posteriori. For example, experimental psychology may discover how certain kinds of reason have causal impacts on people's behaviour—as in the studies I cited in Sect. 2. Or alternatively (and complementarily), a theory of psychodynamics may propose models of how thematic relations between meaningful contents become causes of other contents and mental states (this was Freud's original project, still partially pursued by some heirs to the psychoanalytic tradition). What is missing from an a priori philosophical theory may come as the a posteriori result of a scientific investigation—and in Sect. 2 I suggested that this is what is in fact happening.

4 Concluding Remarks

In this paper I defended the need of a philosophical framework to bridge the traditional divide between reasons and causes in psychiatric explanation. I argued in Sect. 2 that there are two forces that push in the direction of bridging such a divide: multi-factorial causal models of mental disorders (where reasons have a role), and scientific approaches to psychodynamic psychotherapy. I claimed that, if these research paradigms do use reasons as causes, then there should be a philosophical framework

that accommodates that. My suggestion is that some of Davidson's ideas from the debate on the nature of action can be usefully extrapolated and employed, provided that the notion of reason is somehow broadened so to include all meaningful mental states. Of course, my conclusion and suggestions depend on two heavy general assumptions: that psychiatry is at least in part empirically based, and that philosophy itself is continuous with empirical science. The justification of both assumptions far exceeds the limits of this paper.

Notes

1. Relevant references for the debate are Szasz (1960), Laing (1971), Foucault (1961) and Grünbaum (1985).
2. In the next section I qualify the categorization of such conditions as reasons.
3. For interventionism in psychiatry, see Woodward (2008), Campbell (2013) and Murphy (2010).
4. The philosopher and psychiatrist Derek Bolton put forth a positive proposal of how reasons can be causes within the context of psychiatric explanation. His work was published in the 1990s, in the heyday of cognitivism and cognitive-behavioural therapies (1997a, b). Some of the notions he employs, notably the encoding of psychological states by brain states, are tied to the cognitive-representational view of the mind, which has been partially superseded by more complex or nuanced models. This is why I choose not to discuss it here, though I am sympathetic to his motivations and overall conclusion. See Thornton (2007) for objections.
5. Donald Davidson's *Essays on Truth and Interpretation* and *Essays on Action and Events* were included in the syllabus of Eva Picardi's course Philosophy of Language, which I attended in 1995 at the University of Bologna. I graduated two years later with a BA dissertation where I discussed some of Davidson's ideas about communication. Though now my research interests have moved far from that initial focus, it is no exaggeration to say that my encounter with Eva, and with the kind of philosophy she showed us, changed my life completely. From Eva I learnt to appreciate the never-ending beauty of difficult things, and the fragile pleasures of trying to understand them.
6. See D'Oro and Sandis (2013) for a historical overview of the debate between causalism and rationalism in the theory of action, and Glock (2014) for discussion.
7. By 'mental disorder' here I refer to the conditions described in the DSM-5 (APA 2013), of which Depressive Disorders and Schizophrenia are prototypical examples. I bracket here the ongoing discussions about

what counts as a mental disorder, and whether the current nosology is adequate or not.

8. Thomas Szasz (1960) famously argued that, if mental diseases are of the lesion or infection species, they are not mental and that, if they are not, then they are not diseases.
9. More precisely, the study in the example refers to Major Depressive Disorder (see APA 2013 for a characterization).
10. The phrase is that of Wilfrid Sellars (1956), later employed by John McDowell (1996).
11. Or, as I noted earlier, we need an interventionist view of causation, where there are no a priori indications on which factors may influence a given condition (see Campbell 2013).
12. A review of possible biases and errors in this literature is Lilienfeld et al. (2014). See also Luborsky et al. (1999).
13. 'Thematic relations' is an expression used in the literature on psychoanalysis and psychotherapy. See e.g. Grünbaum (2004).
14. There has been a long debate about anomalous monism and its alternatives. For an introduction and a critical position, see Kim (1995, 2000).

References

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. Washington: American Psychiatric Publications.

Boden, J.M., D.M. Fergusson, and L.J. Horwood. 2014. Associations Between Exposure to Stressful Life Events and Alcohol Use Disorder in a Longitudinal Birth Cohort Studied to Age 30. *Drug and Alcohol Dependence* 142: 154–160.

Bolton, D. 1997a. Encoding of Meaning: Deconstructing the Meaning/Causality Distinction. *Philosophy, Psychiatry, & Psychology* 4 (4): 255–267.

Bolton, D. 1997b. Response to the Commentary. *Philosophy, Psychiatry, & Psychology* 4 (4): 273–275.

Campaner, R. 2014. Explanatory Pluralism in Psychiatry: What Are We Pluralists About, and Why? In *New Directions in the Philosophy of Science*, ed. M.C. Galavotti et al., 87–103. Dordrecht: Springer.

Campbell, J. 2007. An Interventionist Approach to Causation in Psychology. In *Causal Learning: Psychology, Philosophy, and Computation*, ed. Alison Gopnik and Laura Schulz, 58–66. Oxford: Oxford University Press.

Campbell, J. 2013. Causation and Mechanisms in Psychiatry. In *Oxford Handbook of Philosophy and Psychiatry*, ed. K.W.M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, et al., 935–949. Oxford: Oxford University Press.

Cuijpers, P., T. Donker, M.M. Weissman, P. Ravitz, and I.A. Cristea. 2016. Interpersonal Psychotherapy for Mental Health Problems: A Comprehensive Meta-analysis. *American Journal of Psychiatry* 173 (7): 680–687.

Cuijpers, P., E.H. Turner, D.C. Mohr, S.G. Hofmann, G. Andersson, M. Berking, and J. Coyne. 2014. Comparison of Psychotherapies for Adult Depression to Pill Placebo Control Groups: A Meta-analysis. *Psychological Medicine* 44 (4): 685–695.

Davidson, D. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy*, 60 (23): 685–700. Reprinted in Davidson, 2001.

Davidson. 1970. Mental Events. In *Experience and Theory*, ed. L. Foster and J.W. Swanson. London: Duckworth. Reprinted in Davidson. 2001. *Essays on Actions and Events: Philosophical Essays*, vol. 1, 207–225. Oxford University Press on Demand.

Davidson, D. 1991. Three Varieties of Knowledge. *Royal Institute of Philosophy. Lecture Series* 30: 153–166.

Davidson, D. 2001. *Essays on Actions and Events: Philosophical Essays*, vol. 1. Oxford University Press on Demand.

D'Oro, G., and C. Sandis. 2013. From Anti-causalism to Causalism and Back: A Century of the Reasons/Causes Debate. In *Reasons and Causes: Causalism and Non-causalism in the Philosophy of Action*, ed. G. D'Oro and C. Sandis, 7–48. New York: Palgrave Macmillan.

Fonagy, P. 2003. Psychoanalysis Today. *World Psychiatry* 2 (2): 73–80.

Fonagy, P. 2015. The Effectiveness of Psychodynamic Psychotherapies: An Update. *World Psychiatry* 14 (2): 137–150.

Foucault, M. 1961. *Folie et déraison: histoire de la folie à l'âge classique*, vol. 169. Paris: Plon.

Glock, H.J. 2014. Reasons for Action: Wittgensteinian and Davidsonian Perspectives in Historical, Meta-philosophical and Philosophical Context. *Nordic Wittgenstein Review* 3 (1): 7–46.

Grünbaum, A. 1985. *The Foundations of Psychoanalysis: A Philosophical Critique*. Berkeley: University of California Press.

Grünbaum, A. 2004. The Hermeneutic Versus the Scientific Conception of Psychoanalysis. In *PsychoAnalysis at the Limit*, ed. J. Mills. New York: SUNY Press.

Insel, T.R., and B.N. Cuthbert. 2015. Brain Disorders? Precisely. *Science* 348 (6234): 499–500.

Jaspers, K. 1923 (repr. 1963). *Allgemeine Psychopathologie*. Berlin: Springer Verlag. English translation by J. Hoenig and M.W. Hamilton, General Psychopathology. Manchester: Manchester University Press.

Kendler, K.S. 2005. Toward a Philosophical Structure for Psychiatry. *American Journal of Psychiatry* 162 (3): 433–440.

Kendler, K.S. 2008. Explanatory Models for Psychiatric Illness. *American Journal of Psychiatry* 165 (6): 695–702.

Kendler, K.S., and J. Campbell. 2009. Interventionist Causal Models in Psychiatry: Repositioning the Mind-Body Problem. *Psychological Medicine* 39 (6): 881–887.

Kendler, K.S., and L.J. Halberstadt. 2013. The Road Not Taken: Life Experiences in Monozygotic Twin Pairs Discordant for Major Depression. *Molecular Psychiatry* 18 (9): 975–984.

Kendler, K.S., C.O. Gardner, and C.A. Prescott. 2002. Toward a Comprehensive Developmental Model for Major Depression in Women. *American Journal of Psychiatry* 159 (7): 1133–1145.

Kendler, K.S., L.M. Karkowski, and C.A. Prescott. 1999. Causal Relationship Between Stressful Life Events and the Onset of Major Depression. *American Journal of Psychiatry* 156 (6): 837–841.

Kim, J. 1995. *Supervenience and the Mind*. Cambridge, UK: Cambridge University Press.

Kim, J. 2000. *Mind in a Physical World*. Cambridge, MA: MIT Press.

Klein, G.S. 1976. *Psychoanalytic Theory*. New York: International Universities Press.

Lacewing, M. 2013. Could Psychoanalysis Be a Science? In *The Oxford Handbook of Philosophy and Psychiatry*, ed. K.W.M. Fulford, M. Davies, R.G.E. Gipps, G. Graham, J.Z. Sadler, G. Stangehellini, and T. Thornton. Oxford, UK: Oxford University Press.

Laing, R.D. 1971. *The Politics of the Family, and Other Essays*, vol. 5. London: Tavistock.

Lilienfeld, S.O., L.A. Ritschel, S.J. Lynn, R.L. Cautin, and R.D. Latzman. 2014. Why Ineffective Psychotherapies Appear to Work: A Taxonomy of Causes of Spurious Therapeutic Effectiveness. *Perspectives on Psychological Science* 9 (4): 355–387.

Luborsky, L., L. Diguer, and D.A. Seligman. 1999. The Researcher's Own Therapy Allegiances: A 'Wild Card' in Comparisons of Treatment Efficacy. *Clinical Psychology: Science and Practice* 6: 95–106.

McDowell, J. 1996. *Mind and World*. Harvard: Harvard University Press.

Murphy, D. 2006. *Psychiatry in the Scientific Image*. Cambridge, MA: MIT Press.

Murphy, D. 2010. Explanation in Psychiatry. *Philosophy Compass* 5 (7): 602–610.

Murphy, D. 2013. The Medical Model and the Philosophy of Science. In *The Oxford Handbook of Philosophy and Psychiatry*, ed. K.W.M. Fulford, M. Davies, R.G.E. Gipps, G. Graham, J.Z. Sadler, G. Stangehellini, and T. Thornton, 966–986. Oxford, UK: Oxford University Press.

Ricoeur, P. 1970. *Freud and Philosophy*. New Haven: Yale University Press.

Sellars, W. 1956. Empiricism and the Philosophy of Mind. In *Minnesota Studies in the Philosophy of Science*, vol. 1, ed. Feigl and Scriven. Minneapolis: University of Minnesota Press.

Shea, N. 2013. Neural Mechanisms of Decision-Making and the Personal Level. In *The Oxford Handbook of Philosophy and Psychiatry*, ed. K.W.M. Fulford, M. Davies, R.G.T. Gipps, G. Graham, J.Z. Sadler, G. Stanghellini, and T. Thornton, 1063–1082. Oxford, UK: Oxford University Press.

Stanghellini, G., and T. Fuchs. 2013. *One Century of Karl Jaspers' General Psychopathology*. Oxford, UK: Oxford University Press.

Sullivan, J.A. 2016. Models of Mental Illness. In *The Routledge Companion to the Philosophy of Medicine*, ed. H. Kincaid, J. Simon, and M. Solomon, 455–464. New York: Routledge.

Szasz, T.S. 1960. The Myth of Mental Illness. *American Psychologist* 15 (2): 113–118.

Tennant, C. 2002. Life Events, Stress and Depression: A Review of Recent Findings. *Australian and New Zealand Journal of Psychiatry* 36 (2): 173–182.

Thornton, T. 2007. *Essential Philosophy of Psychiatry*. Oxford: Oxford University Press.

Woodward, J. 2008. Cause and Explanation in Psychiatry. In *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*, ed. Kenneth S. Kendler and Josef Parnas, 132–184. Balitmore: Johns Hopkins University Press.

Yalowitz, S. 2014. Anomalous Monism. In *Stanford Encyclopaedia of Philosophy*, ed. Edward N. Zalta. https://plato.stanford.edu/archives/win2014/entries/anomalous-monism/.

The Doxastic Zoo

Pascal Engel

1 Introduction: The Menagerie of Belief

The task of classifying the species of belief seems to elude us. The very idea of asking what one believes at a given time, or how many beliefs a person has, or even whether someone believes rather than is in another state with respect to a given content, seems not to make real sense. Nevertheless we ascribe beliefs to ourselves and to others, we talk of acquiring or losing them, and we give them a number of properties, such as being silly or wise, rational or not, justified or not. Some philosophers tell us that belief is such a simple and unanalysable notion that it cannot be defined.[1] Others, like Wittgenstein, suggest that it is so diverse that

Versions of this article were presented in 2012 at the SIFA conference in Alghero, and in the Brasilian conference on analytic philosophy in Fortaleza. I thank my friends and participants for their invitations and comments.

The nature of belief and judgment are topics on which Eva Picardi wrote in various ways. I learnt a lot from her. It has been a great luck for me to come to know her.

P. Engel (✉)
EHESS-Ecole des hautes études en sciences sociales, Centre de recherche sur les arts et le langage (CRAL), Paris, France
e-mail: pascal.engel@ehess.fr

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_14

the best that we can say is that it consists in a motley of psychological states, which bear with each other only family resemblances. But, as J.N. Findlay (1962: 94) remarked, this kind of answer is no answer, since it attempts to dispel our sense of unity inadequately brought out in a range of cases, without showing this sense to be ill founded. "Belief" may not be defined, but it indicates a place in the map of the mind, and we can at least indicate the role of this state among other states. I cannot here give a full picture of the family tree of belief, but at least hope to give a profile of this animal within the doxastic zoo.

Part of the difficulty comes from the fact that belief does not stand on its own among other mental states. Many psychological attitudes are related to belief, in the sense that they are often species of belief, or involve the possession of belief. In the most general sense, belief involves assent, or holding-true a proposition. Beliefs vary depending on the nature of the expression or vehicle. Believing may be a mere disposition to act, with no conscious or verbal expression. It may be merely tacit. It may be an inner conscious assent in the mind without verbal expression. It may be expressed in an act of linguistic assertion. It may be occurrent or long-standing. It may vary in degrees, ranging from merely subjective opinion, as a conjecture or hypothesis, a mere guessing, or a firm conviction. These degrees may depend on the weight of the evidence that subjects claim to have, in which case the beliefs are held with more or less confidence or certainty. But can these degrees of confidence be quantified? Are there degrees of beliefs and credences? Beliefs can also be associated to voluntary acts, as in judgments, decisions as acceptances, commitments, takings for granted, postulations and presupposition, anticipations or presumptions. They can be more or less firm, such as doubts, conjectures and convictions. There are attitudes related to beliefs or based on beliefs, although distinct from them, such as noticings, perceivings, intentions, regrets, hopes, expectations, or indeed judging. All of these attitudes may be related to different epistemic sources and variously to knowledge, to different emotions, they can be more or less firm, more or less long-lived. Some may be so deeply entrenched, such as "hinge propositions" that many philosophers do not take them as beliefs. All of these kinds of belief-like attitude I shall call beliefs' *cousins*. In what sense are they beliefs, and if they are, how do they relate to this genus? What is the shape of the family, and how remote are the cousins? Such questions have resurfaced very often in the history of philosophy, but they have become even more complex when contemporary

epistemology and psychology have investigated various kinds of mental states, which one might call belief's "strange bedfellows" (Bayne and Hattiangadi 2013). First, there are the "tacit beliefs" (Lycan 1985), which that one has not never thought about but that one is disposed to believe when told, such as "Kant's earlobe is saller than the Sea of Tranquility". There are "subdoxastic states" (Stich 1978), which are informations, and perhaps representations, processed in our perceptual or memory systems when we perform various cognitive tasks (such as sketches in visual processing, or "mental models" in reasoning) but which are not properly believed. There are delusions, or pathological beliefs, such as Capgras'syndrom, where people "believe" that their friends or family have been replaced by impostors, or Cotard's delusion, where people "believe" that they are dead (Davies and Stone 1992; Bayne and Pacherie 2005). There are the "feelings of knowing" (Koriat 2005), feelings of familiarity, of "déjà-vu" or of strangeness. Finally, there are the so-called *aliefs* (Gendler 2008), which include phobias, and various emotion-induced feelings and representations, and *biases* which are supposed to influence our attitudes and our reasoning about various subject matters (Brownstein and Saul 2016). In each of these cases, there is a debate as to whether such states are beliefs at all. These issues loom large in contemporary philosophy of mind and epistemology, and I cannot hope to solve these problems case by case, or to go into the widely discussed topics of the relations between belief an emotion, belief and inference, belief and mental representation, belief and the will, which stand behind these issues.[2] Nor do I hope to produce a full taxonomy of belief-attitudes or of the doxastic realm. My main aim here is to discuss the methodology of the philosophy of belief, one of the main problems of which is to establish a classification of beliefs and other related attitudes. I shall try to give a criterion to individuate beliefs and locate them with respect to other epistemic attitudes.

2 What's in a Belief, That Ink May Character?

Let us try to give the usual profile of belief. This should not be a definition, but a set of characteristics which are typically ascribed to beliefs. These are the following:

i. Representational content: Beliefs are attitudes with a (linguistic or not) content
ii. Direction of fit: beliefs "aim at truth" and are supposed to be adapted to the world

iii. Causal or motivational profile: beliefs are dispositions to act as if their contents were true
iv. Reason and evidence: beliefs are supposed to have reasons, which consist in evidence
v. Involuntariness: belief are not (normally) under the control of the will
vi. Degree: beliefs have degrees, in being more or less certain
vii. Inferential integration: beliefs are supposed to be inferentially coherent
viii. Immediate access: most beliefs can be in principle consciously accessed
ix. Context independence: believing is not relative to context
x. Emotional impact: beliefs have at least some causal relations with emotions
xi. Phenomenology: some beliefs at least have a certain phenomenology.

These traits have to be spelled out through various theories of belief, which conflict on crucial points. Thus dispositional of functionalist theories emphasise the dispositional and causal features of beliefs, which entail that they are not necessarily conscious and that they are necessarily action-oriented, whereas "Cartesian" theories stress this very feature and say that the practical properties of belief are not primary. Some philosophers deny that beliefs have degrees, while others take the latter to be degrees of subjective probability. Some take belief contents to be semantic, meaning that they are necessarily linguistic. Some hold that beliefs can be voluntary. Not all theorists agree that beliefs have a specific connection to emotions and have a phenomenology. In spite of these differences, everyone agrees that a least some beliefs possess some of these features. Moreover these features apply mostly to *rational* beliefs, those which are supposed to be true, consciously accessible, held for reasons, and inferentially coherent, whereas the strange bedfellows are most of the time *irrational* beliefs, if they are beliefs at all. Some, perhaps most, beliefs arise from causes which have nothing to do with the evidence or reasons that one has for them. Some people are prone to wishful thinking and do not care for having true beliefs. Most religious beliefs are not held for reasons having to do with evidence or coherence. This raises the question whether these eight features can really characterize beliefs in general. As soon as one raises such questions, one is tempted to locate

the characteristics under another heading: acceptances, commitments, credences, etc., which in turn differ from credences and dispositional beliefs. Are these *modes* of belief or distinctive *kinds* of doxastic states? A mere taxonomy, however, cannot solve a philosophical problem.

Our problem is to individuate beliefs and to locate them within the broader territory of propositional attitudes, and most specifically of epistemic attitudes. The latter are attitudes aimed at truth and that depend upon evidence. In this respect, not only belief, but also perception, judgment, inference, supposition, some cases of imagination, and knowledge itself can count as epistemic attitudes. But there is a problem about the frontiers of the epistemic domain. Not only is belief sometimes hard to distinguish from other attitudes, but it is also central to other psychological states, which often presuppose the possession of beliefs. Perceiving as an experience is not believing, but it often depends on beliefs and leads to beliefs. To be surprised is to discover that P when one believed that not P. Imagining is not believing but is hard to understand without the existence of belief. Can one have desires without having beliefs about what one desires? Most attitudes and emotions presuppose belief: if one regrets something, one believes that it is the case and desires that it had not been the case,, if one is angry at someone, one believes that someone did something offensive,, etc. Belief is as much characterized by its causal and logical links with other attitudes as it is characterisable as a single and isolated state.

So why not characterize belief in functional terms instead of trying to define belief, by describing it as the state which is the conjunction of the platitudes (i)–(viii), conjoined in a Ramsey sentence? It would look like this:

(∃ S) such that (∀ X) (∀ p)

i. X is in some state which represents p
ii. If X believes that p x is disposed to act as if p
iii. X believes that p on the basis of evidence
iv. If X believes that p, p is not under the control of X's will
v. if X believes that p and q can be inferred from p X is disposed to believe that q
vi. If x believes that p, x is disposed to be in some emotional state
vii. X believes that (s)he believes p
viii. If X believes that p, there is a certain phenomenology.

These are the most common features of belief, and its cousins and bedfellows instantiate them in most, but not all cases. The minimal core of beliefhood would be the representational property (i) and causal functional property (ii): belief is a kind of contentful state which has certain causes and certain effects. This very wide characterization would certainly allow us to include among beliefs the so-called "aliefs", such as the kind of vertigo which one often experiences in elevators with glass floors, epistemic feelings such as *déjà vu*, as well as attitudes like acceptance, where one can take a proposition for granted without believing it. Such states share some of the characteristics of the general profile (i)–(viii), but not all: they involve representations, have a causal profile, a phenomenology and an emotional tone, but they are not in general associated with a reflective mode or with a rational and evidential role. But this functionalist strategy yields too loose a categorization. Virtually every state which involves a belief, either as a cause or as a consequence, becomes a belief. Thus attitudes such as hope, regret and a number of emotions also involve the minimal core of representation *cum* causal profile, but it would be wrong to include these within beliefs. If, however, we cut our doxastic cake too narrowly and exclude from the doxastic club all but the states which conform to the (i)–(viii) features, we risk chauvinism. Moreover, we should not multiply kinds of beliefs beyond necessity. Consider cases of negligence or forgetfulness. I take my broken watch for repair at the watchmaker. A few moments after leaving the shop, I want to know what time it is, and consult my empty wrist: I forgot that I had left my watch for repair. Do I believe that my watch is on my wrist or not? Maloney (1990) suggests that there are two kinds of beliefs, those which are dispositional and action-oriented and not necessarily conscious ("A-Beliefs") and those which are consciously entertained or accessible ("C-Beliefs"). A very simple-minded reaction would be to say that I believed that my watch was on repair, but forgot it, hence temporarily ceased to believe it, or revised my belief. Is it necessary here to postulate two kinds of beliefs? It is always costly to posit too many entities, but such examples are enough to show that among the features (i)–(viii) one can privilege the dispositional and motivational ones, or those which stress the conscious, linguistically manifestable aspects. The philosophy of belief in this sense is not exceptional in leading to an opposition between the functionalist third-person perspective on the mind and the first-person "Cartesian" perspective. If we privilege the former, we take belief to be mostly a dispositional/causal profile, and take accessible beliefs to be merely epiphenomenal. Thus we are able to include many states which, like aliefs,

delusions and epistemic feelings, are unconscious or semi-conscious, but we run the risk of excluding those beliefs which are the products of our commitments and of our "judgment sensitive attitudes" (Scanlon 1998).[3] Alternatively, we could envisage a third category: Dennett (1978) talks of "opinions", Sperber (1985) talks of "quasi-beliefs", and Schwitzgebel (2001) talks of "in-between believings". These might be located in between the conscious and dispositional ones. Dennett has a distinction between beliefs as well entrenched dispositions, and "opinions" which are merely verbal and short-lived assents to a sentence. Of many of the contents which float in our minds we are unsure whether they are beliefs or not, in the sense of being apt to be expressed as assertions. As Bernard Williams says: "It is far from being true that every thought swimming around in one's mind is already the content of a belief as opposed to some other mental state such as a guess, a fancy, or (very importantly) a wish. […] in many …cases, it is not merely that we do not know what we believe (though this is of course often true), but that a given content has not come to be a belief at all" (Williams 2002: 82).

So the category of "in between believings" is very large indeed. It contains all our short-lived assents, our tacit beliefs and it can include aliefs and other epistemic feelings or emotions. The problem with the strategy of dubbing these beliefs "in-between" is not that such phenomena aren't real, but that it is difficult to suppose that there is a specific category of intermediate believings. Aren't in one sense *all* believings "in between" believings? But if we make this move, we simply rehearse the functionalist strategy: for the intermediary beliefs are those which possess all the features of belief which are action-oriented. How to draw the line? Compare with colours (Zimmerman 2007). We can improve our colour scheme by locating an intermediary colour, say purple, between blue and indigo. But we cannot improve the scheme by adding a colour dubbed "in between blue and red". If we apply this strategy to belief we are led to ask whether the state of mind a person is in is a belief. We could improve the scheme by showing that there are genuine psychological states, presumably physical, which underpin the so-called "in-between believings". The notion of "in-between believing" at best deflates our attempt to taxonomise the varieties of beliefs, and at worst has eliminativist consequences. For it acknowledges that most of what we call "belief" does not fall within one category or another, and is at best a mere instrument of prediction at the service of an "intentional stance". The only way in which we could remove this fuzziness is by locating the neurophysical states responsible for the entities of our folk

psychology. We would encounter the same sort of problems if we used a distinct scheme, such as the one inspired by the famous "two minds" or " two systems" view of cognition advocated (Kahneman 2012; Evans and Frankish 2012) Should we say that there are two kinds of beliefs, those which are the products of System 1 and those which are the products of system 2? The same problem would arise if we managed to locate beliefs within three familiar divisions of the brain: procedural with the cortex cerebellum and striatum, declarative with the hippocampus, emotional with the amygdala. Should we say that dispositional beliefs are in the first, conscious and asserted beliefs in the second, and that all the strange bedfellows, such as aliefs and feelings, are in the third? Or should we say that there is only one kind, belief, which spreads over the three brain systems? Such divisions do not help.

We find similar difficulties when trying to add a new kind of animal to our doxastic zoo, *credences*. Credences are not beliefs, in the sense of full beliefs, but degrees of belief or partial beliefs, the states in which we are when we are not fully confident that a certain proposition is true. According the Bayesians, these degrees are degrees of probability. Notice that, for Bayesians, all beliefs are *essentially* probabilistic. Probability is not in the content of some beliefs, but in the attitude of belief itself. It is not that we believe that something is, say, 0.5, or 0.7 probable. It is that we have a 0.5 or 0.7 credence in it happening. Every belief is thus a probabilistic belief (even if the credence happens to be 1 or 0). But the same problem as with in-between belief arises: should we say that full belief and full disbelief are a degree of belief 1 and 0, that half beliefs are not beliefs? Or should we say that there are full or all-out beliefs which do not have degrees on the one hand and credences or degrees of belief on the other hand (Kaplan 1996)? These difficulties are well known. But for our purposes here, which have to do with descriptive and psychological adequacy, the main difficulty is that Bayesianism is implausible if it amounts to the reductive claim that *all beliefs*, and all the members of the doxastic family, have degrees which are degrees of probability. Bayesianism may be defended as a theory of what rational belief ought to be, but not as a descriptive theory. A lot of psychological work shows that humans are not good Bayesians, and that full beliefs cannot be reduced to partial beliefs (Holton 2014). If we want to preserve the idea that credences are a specific kind of belief, we have to say that they refer to one kind of belief only, perhaps a certain kind of dispositional belief, whereas our flat-out or full beliefs constitute another kind.[4]

These considerations favor a pluralistic or layered view of belief and its cousins and bedfellows, each forming a specific kind. Chauvinism, which reduces belief to one privileged dimension would leave out too much. The deflationary functionalist strategy leads to the view that there are no real differences between beliefs proper and in-between beliefs. It unifies the domain of belief-attitudes at the cost of blurring the categories: we lack any proper way of individuating beliefs and distinguishing them from their subspecies. Nor can we classify beliefs into dualistic schemes, such as those which have a definite dispositional profile in opposition to those which are conscious or accessible to reflection, or between full beliefs and credences. For these divisions are not exhaustive. We need a better criterion.

3 A Normative Account of Belief

The account of belief I propose belongs is "essentialist" (Hazlett 2013): it takes belief to have a distinctive nature, which allows us to set this attitude apart from other doxastic attitudes and from the bedfellows. The main argument in favour of this view is that the list of platitudes (i)–(viii) leaves out the normative dimension of belief. We evaluate beliefs as rational or irrational, as justified or not, as well confirmed or not, as good or bad to have, as obligatory or not. Mere feelings, experiences, dispositions, or mental episodes, are not evaluated as good or bad: one can only say that one has them, or that one tends to have them in many circumstances. The same would be true about sentiments in the moral domain: if these were mere psychological states of approval, we could not say that they are good or bad, rational or not, evaluable as right or wrong. The same applies to belief. It involves a normative essence. States that conform to one or the other of the features (i)–(viii), may resemble belief and sometimes are called "beliefs", but if they are merely defined in causal terms, they will be called "beliefs" only by courtesy. Without the normative dimension, a mental state, however close to belief, is not a belief. This is not to deny that belief has a causal and natural profile. Beliefs can have all sorts or causes and effects and are psychological states in the mind. But qua beliefs, they are evaluable. A state which one could not characterise *both* in causal and in normative terms would not be a belief.[5]

What kind of normative properties are distinctive of belief? Evaluative or axiological ones, such as good or bad, valuable or not valuable? Or deontic ones, such as right or wrong, correct or

incorrect? Some philosophers prefer the concept of *reason*. There are a number of debates on this issue. One way to avoid these debates is to use a generic notion, that of *correctness*. The normative properties relevant for attitudes are *correctness properties*, and that such properties involve the presence of specific *norms*. Belief is governed by a norm of truth: *a belief is correct if and only if it is true*. The exact form of the norm, and the kind of guidance which it is supposed to give us is a matter of discussion.[6] This norm is unique to belief, and sorts it out from the other attitudes and belief-like states. I defend two claims:

i. All attitudes have specific correctness conditions
ii. Belief is the only attitude whose correctness condition is truth.

The correctness condition of an attitude is the condition in which it is appropriate or "fit" to its object. This is related to the familiar idea of the "directions of fit" of attitudes, but it is distinct: directions of fit are supposed to be either mind-to-world (epistemic attitudes) or world-to-mind (conative attitudes), but within attitudes of each kind there can be distinct kinds of objects to which they are fit. Each kind of attitude has a certain typical object, to which it is supposed to adapt or which is adapted to it. This object is neither the *type* of intentional or propositional content of the attitude nor its *token* content (*e.g.* beliefs as a type have an intentional content, which in a particular occasion may be such or such a proposition). Intentional content is not individuative of the attitudes, for the same intentional content can be the object of different attitudes. Thus I can believe that Trump is president, desire that Trump is president, be horrified that Trump is president, regret that Trump is president, etc. And attitudes are not individuated by their intentional content since a single attitude, say desire, can have many different intentional contents. The correctness thesis says that there are types of content which are appropriate to each attitude. It is better formulated in terms of the traditional distinction between the *formal object* of an attitude and its intentional object.[7] The formal object of an attitude is the kind of object to which typically the attitude is directed to. It contrasts with its particular object, which is the specific object of the attitude in a given circumstance. This idea is more familiar in the case of emotions. The formal objects of emotions are in general evaluative properties. Thus the formal object of fear is what is *fearable*, the formal object of love is what is *lovable*, etc. The particular object of fear could be on a given

occasion an earthquake, a monster or a spider. The particular object of love could be Juliet or Romeo or my cat Felix. If we extend this to other attitudes we would say that the formal object of desire is what is desirable, and that the particular objects of desire could be a beautiful person or a big amount of money. The objects may be individuals, propositions or states of affairs. Thus one can be afraid that an earthquake arrives, or that a spider is on the wall, or desire that Juliet loves one. The obvious problem with this proposal is that it looks trivial or false, since there does not seem to be a class of things which objectively have these evaluative properties. Some spiders are worthy of fear, others not, some people or animals are worthy of love, others not. But let us leave that difficulty aside, and let us see how it can apply to epistemic attitudes. The formal object of belief is not the believable, since any proposition can be in principle believed (except perhaps paradoxical ones, such as "This proposition is believed by no one"). Rather, it is truth, and truth makes the attitude properly *epistemic*. The fact that truth is the formal object of an attitude should not be confused with the fact that a number of attitudes which are not epistemic can be expressed as having a propositional content. Thus conative attitudes have this feature—to desire that P is often to desire that P *is true*, to hope that P is to hope that P *is true*—and a number of attitudes or emotions such as imagining, fearing, can take propositional complements. Neither is the fact that the formal object of an epistemic attitude is truth to be confused with its direction of fit. Why? Because the fact that one can construe attitude verbs with a propositional complement, hence with contents susceptible of being true or false, does not imply that the direction of fit and the formal object of these attitudes is truth. Although imagining involves a relation to propositions—one can imagine that one is an Oxford don, or that China invades the U.S—is it not clear that the *aim* of imagining is truth, *i.e.*, that one imagines that P *with the aim of accepting P only if it is true* (Velleman 2000). One can perfectly imagine or hope situations one does not believe or imagine to be true: the point of imagination and of hope is to be able to do so. The potentially propositional nature of the content of such attitudes entails that the predicate *is true* can automatically be affixed to these contents. But it does not entail that their formal object is truth. The formal object of an attitude is the object to which the attitude is "fitted" in the sense of being the *appropriate* or *correct* one, for this attitude. Correctness is the feature an attitude has when it is fit to its formal object. But the latter is not the direction of fit. We should distinguish the correctness condition from the *satisfaction* condition. In the

case of belief and other epistemic attitudes, the satisfaction condition is that the belief is true or that the corresponding state of affairs obtains. It is, as Mulligan (2007) puts it, *because* the belief is true that it is satisfied. In the case of a desire, or a hope, or a wish, the satisfaction is that desire, the hope or the wish be satisfied or realized. In the case of conjectures, these are satisfied when they are probable. But the belief, or the desire, may be satisfied by sheer luck: for instance my belief that 225 + 333 = 558 may be satisfied by pure guess. The correctness condition says more: it says that if one has a belief P it *ought to be true* if it is to fulfill it satisfaction condition. So the satisfaction condition involves necessarily a normative term. It does not matter at this point whether this term is a deontic or an evaluative one, and what kind of guidance it can give to believers.[8] But it has to capture the idea that the attitude *requires* a certain kind of object and that the object *has* to be appropriate to the kind of attitude. So this formal object is a normative property.

Contrast now belief with different attitudes. Some attitudes, like conjectures, hypotheses, suppositions, assumptions[9] and possibly guesses, do not have truth as their formal object and as what fulfills their correctness condition: their formal object is what is probable. This is very plausibly the case for credences, or states of confidence. Conative attitudes have distinctive formal objects, which are plausibly the good, or the appearance (the guise) of the good.[10] Interrogative attitudes such as questions have as formal object what is question-worthy or questionable, doubts have as formal objects what is doubt-worthy or doubtable. Emotions have value properties as formal objects, and are correct when these objects are appropriate (Teroni 2007). I shall not here attempt to offer a list. Let us consider now belief. Why is it special among the attitudes? The formal object of belief, as an epistemic attitude, is truth. Other epistemic attitudes, such as perceivings, noticings, discoverings, surprises, some kinds of suppositions and hypotheses, and indeed knowledge can have truth as their formal object. Surprises are plausibly epistemic emotions, which arise from a contrast between what one believes to be true and what one expected to be true. But the correctness of a belief is specific. It does not consist simply in the fact that it is "aimed at truth" in the sense that it has a direction of fit. It consists in the fact that, as Williams says, it is a "fatal" flaw for a belief to be false: something is wrong, in the sense that the belief cannot be held or maintained if it is show to be false. Belief is the only attitude which is correct only if true. The correctness condition involves the nature of the success of the

attitude. A belief is successful if true, and it fails if false. In a sense truth as the correctness or success condition seems to apply as well to suppositions, guesses, or surprises. A guess which is true is indeed successful, as well as a supposition. But one can make a good guess which does not hit truth, and which is only almost true. And one can suppose successfully without one's supposition being true. One can be surprised at something which turns out to be false. The success of a guess or of a supposition does not depend crucially on its being true, but on the way in which evidence has been collected. Indeed the best suppositions are those which are true, but a supposition which turns out to be false is not necessarily wrong. Imagining has an intentional object, and to imagine that P is to imagine that P is true, but the success of an act of imagination does not depend on it delivering true outputs. The correctness condition of supposing consists in respecting a different norm from truth: evidence. Evidence is indeed evidence for truth, but supposing, can be successful even when the truth target is not hit. Similarly for guessing: a true guess is indeed successful, but it is not simply correct when true. Thus in a quiz it is as important to guess quickly as to guess truly (Owens 2003). Belief, in contrast, is correct if and only if true. There is no room for balancing the aim of truth, or the correctness condition of being true, with other aims.[11] This sets belief apart from other epistemic attitudes. A question might arise, however, about knowledge. Isn't its correctness condition truth? Indeed it is, but its correctness condition is *automatically* fulfilled: knowledge is factive, since it entails the truth of the content. This is not the case for belief. Thus, knowledge and belief share their correctness condition, although in the case of knowledge this condition cannot fail to obtain

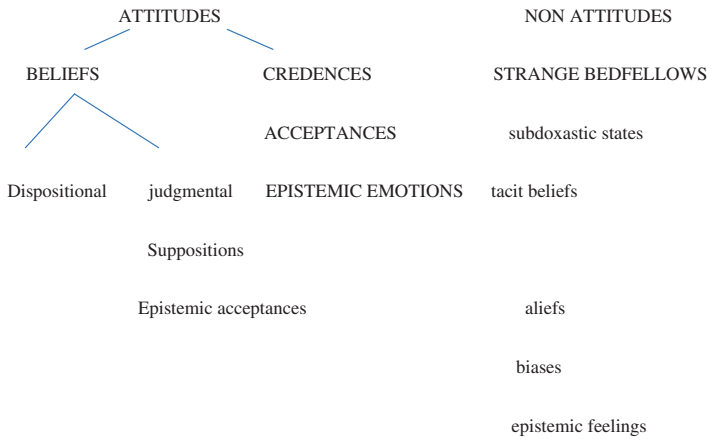
4 The Doxastic Taxonomy Revisited

The taxonomy I propose does not amount to a theory of belief, but it presupposes it. In the first place, I reject eliminativism: I take most states admitted by common sense psychology to be real. I am less sure, however, that the states which philosophers have designated as "in-between" beliefs or "aliefs" belong to the doxastic zoo, although these may be close of some species (for alief, close to beliefs and to emotions). Second, I agree with the functionalists, pragmatists and dispositionalists about belief that belief is at least a disposition to act and is in part individuated by its causal role. Belief is certainly a complex, multitrack

disposition (Ryle 1949). But this does not place it apart from other dispositional states, such as epistemic feelings and epistemic emotions, such as surprise, wonder, or feelings of familiarity. Nor does the dispositional theory allow us to differentiate belief from such strange bedfellows as "aliefs", tacit beliefs, and biases on the one hand, for all these states have a behavioral and dispositional profile. The dispositionalist view does not allow us to distinguish belief from the higher-order judgmental attitudes such as suppositions, conjectures, acceptances, commitments, because these states are clearly not simply dispositions to act. The criterion I have proposed is: *belief is the only attitude whose correctness condition is truth.* This separates off belief from states which are not attitudes and most of its strange bedfellows. There are no correctness conditions for epistemic feelings, such as feelings of familiarity, feelings of knowing or feelings of *déjà vu* (Dokic 2012): these can be successful, and make us able to locate in memory the content we try to retrieve, but they can also fail to deliver any content. There is no rule or norm for their success, or any condition which could constitute what it is for such states to be correct. They are not *attitudes* at all, since it is not clear that they have a propositional content. They certainly are based on some kind of evidence, but it is always unclear where this evidence lies for the agent, and to what use it can be put to infer anything. It is not even clear what their contents are, and in this sense they do not have formal objects. They do not have reasons, and cannot be held for a reason. Their content in some cases is exhausted by their phenomenology, and we can therefore be skeptical that there be a "cognitive phenomenology" in their case (Bayne and Montague 2012). One may doubt that aliefs and tacit beliefs be kinds of belief, in spite of the fact that they may involve or be related to beliefs (*ditto* for biases). I would also reject delusions as clear cases of beliefs, although I would be prepared to take them into an intermediary status (Bayne and Pacherie 2005) because they do not have correctness conditions. By disqualifying most of the bedfellows as beliefs and as attitudes and by excluding them from the doxastic club, it seems that I take as the criterion of an attitude the fact that these are under the agent's control and that one can be responsible of the attitude that one takes. This is true of some attitudes, those of the judgmental kind, but not of all attitudes. Emotions, by most criteria, are attitudes which have correctness conditions and formal objects. But in most cases we do not control our emotions and are not responsible for them. Epistemic emotions, such as surprise and wonder, are not attitudes for which we are responsible, or which involve any kind of agency. But like all emotions, including non-epistemic ones, they

are based on beliefs which are their cognitive bases, even though they are not beliefs, because they do not have truth as their correctness condition. Nor can we include *hinges* within beliefs. They do not have truth and correctness conditions, are not subject to epistemic norms, and in this sense are not attitudes at all (Engel 2015).

In driving a wedge between beliefs and their strange bedfellows, it seems that I exclude *credences* from beliefhood. Bayesians would disagree. Actually it is arguable that credences share a lot with beliefs in that they have correctness conditions, obey norms (Buchak 2014) and are sensitive to evidence and to principles of rationality. Thus, they are not strange bedfellows, but close cousins. Yet, by definition their correctness condition is not truth, but high probability. According to the normativity and correctness criterion proposed here, only judgmental attitudes or attitudes which are judgement-sensitive are beliefs properly so called: judgments, epistemic acceptances and commitments all qualify, because they all can be taken as premises for one's reasoning. These may or may not be voluntary in a strong sense: for some acceptances may be only implicit. But they are all categorical or all out, and do not have degrees. They are all subject to a correctness norm, which is truth. Some acceptances are pragmatic, contextual, and voluntary (Cohen 1992). In this sense they are not beliefs. But some kinds of acceptance, those which have an epistemic aim, such as hypothesizing, or taking certain things for granted, are clearly judgmental attitudes. On the basis of these classifications, one could try to give the following preliminary taxonomy:



A lot more should be said to refine these categories. To classify a state under the heading of *belief* implies that (i) it has a causal role, (ii) is non voluntary (iii) is sensitive to reasons and to evidence, and most of all is an attitude which has truth as its correctness conditions. Some beliefs can be purely dispositional in the sense that they are revealed in behavior (including verbal) and are not reflectively conscious. Others are judgmental, in the sense that they are the object of a conscious and all-out assent (not necessarily verbal). They can be voluntary, not because they can be willed for non-epistemic reasons (such as pleasure or comfort), but because in holding them one commits oneself to them, and to the inferences to which they lead. Some of them are parts of rational plans and strategies, such as epistemic acceptances. Credences enjoy an intermediary status, being both very close to dispositional belief by being in large part non reflective, attitudinal, and normative. But their correctness condition, probability, is not truth, hence they are not beliefs. Pragmatic acceptances are attitudes with correctness conditions, aimed not at truth, but at utility. There can be also intermediary epistemic attitudes, such as conjectures, suppositions, expectations, presumptions and hypotheses, with correctness conditions, but not clearly truth oriented. Hypotheses are closer to epistemic acceptances, and presumptions closer to dispositional beliefs. But they are not judgmental in the sense that a thinker does not commit himself to them as he usually does in the case of judgments. All other states fail, in one way or another, to be attitudes, to have correctness conditions, and to be evidence oriented. They are all epistemically loaded, as they involve the processing of information. They all have a causal and behavioral profile, but they do not involve the kind of assent, reason and normative basis that goes with belief. They can, phrase, constitute "seeds of knowledge", but they are not in the business of knowing (Dokic 2012).

I have not included knowledge among these attitudes because knowledge is, strictly speaking, neither a species of belief nor an attitude, although it entails belief. Knowledge entails truth, but it would be wrong to say that truth is its correctness condition. Knowledge is not only justified or warranted belief, but justified and warranted belief entailing truth. It is actually the norm of belief, what belief aspires to. Knowledge is the ideal of belief., I shall content myself to indicate how it relates to the nature of belief, when we compare it to other attitudes. Our attitudes, such as belief, desire, hope, etc., are associated to various

presuppositions about the kinds of judgements, inferential relations, causal profiles, and strategies that they involve. These presuppositions are in many ways contingent and can vary, depending on social, historical, or cultural factors. They are in various ways tied to our nature, and can depend on all kinds of individuals' circumstances: some people are afraid of spiders, many believe and desire weird things. These contingencies, however, do not affect the essence of belief. Each attitude has an ideal profile, one which it ought to have. It is particularly the case for belief: it is an attitude which is associated, implicitly or not, in a believer, to what he takes himself to believe, to what he considers that he ought to believe. I have called this feature here "correctness condition", but one could call it the ideal of belief. This ideal gives to belief the primacy within the doxastic zoo. Ideals are limits, and one can be more or less far from the ideal. In the case of belief with respect to other epistemic and doxastic attitudes, the purest form of the ideal is *judgement* (Findlay 1954; Sosa 2015). The closer a belief kind of state is to this ideal, the closer it is to the core of believing. The farther it is from this ideal, the less it is part of the doxastic realm.

Notes

1. Hume, *Inquiry* (II) Mill (1843, book I, Chapter v, §1: 88), Brentano (1874, I, 64) and Stout (1896: 99) take belief and judgment to be primitive and undefinable. Cf. van der Schaar (2013).
2. I have often discussed these issues, e.g. in Engel (1998, 1999, 2000, 2013a, b, c).
3. A number of writers have emphasized this dimension of belief: Cohen (1992), Hieronymi (2008), Gilbert (2013), Holton (2014).
4. Leitgeb (2017) defends this "independence option".
5. *Contra* see Bennett (1990) and Papineau's (2013).
6. See Bykvist and Hattiangadi (2007), Mc Hugh (2012), Engel (2013c).
7. The distinction is a medieval one, taken up by Brentano and his disciples, reintroduced by Kenny (1963). See also De Sousa (2005), Teroni (2007), Mulligan (2007).
8. See Glüer and Wikforss (2009), Steglish-Pedersen (2010).
9. On assumptions see Meinong (1902).
10. Velleman (2000), Deonna and Lauria (2017).
11. Engel (2013a, b), *contra*, Reisner (2008).

References

Bayne, T., and A. Hattiangadi. 2013. Belief's Strange Bedfellows. In *New Essays on Belief*, ed. N. Nottelman. London: Palgrave.

Bayne, T., and M. Montague (eds.). 2012. *Cognitive Phenomenology*. Oxford: Oxford University Press.

Bayne, T., and E. Pacherie. 2005. In Defense of the Doxastic Conception of Delusions. *Mind and Language* 20 (2): 163–188.

Bennett, J. 1990. Why is Belief Involuntary? *Analysis* 50 (2): 87–107.

Brentano, F. 1874 (1971). *Psychologie vom empirischen Standpunkt*. Hamburg: Felix Meiner.

Brownstein, M., and J. Saul. 2016. *Implicit Biases and Philosophy, 1: Metaphysics and Epistemology*. Oxford: Oxford University Press.

Buchak, L. 2014. Beliefs, Credences and Noms. *Philosophical Studies* 169 (2): 285–311.

Bykvist, K., and A. Hattiangadi. 2007. Does Thought Imply Ought? *Analysis* 67 (296): 277–285.

Cohen, J. 1992. *An Essay on Belief and Acceptance*. Oxford: Oxford University Press.

Davies, M., and T. Stone (eds.). 1992. *Pathologies of Belief*. Oxford: Blackwell.

Dennett, D. 1978. How to Change Your Mind. In *Brainstorms*. MA: Bardford Books.

Deonna, J., and F. Lauria (eds.). 2017. *The Nature of Desire*. Oxford: Oxford University Press.

De Sousa, R. 2005. Emotional Truth. *Proceedings of the Aristotelian Society* 76 (76): 265–275.

Dokic, J. 2012. Seeds of Knowledge. In *The Foundations of Metacognition*, ed. M. Beran, J. Brandl, J. Perner, and J. Proust, 302–321. Oxford: Oxford University Press.

Engel, P. 1998. Believing, Accepting, and Holding True. *Philosophical Explorations* 1 (2): 140–151.

Engel, P. 1999. Dispositional Belief, Assent and Acceptance. *Dialectica* 53 (3–4): 211–226.

Engel, P. (ed.). 2000. *Believing and Accepting*. Dordrecht: Reidel.

Engel, P. 2013a. Trust and the Doxastic Family. *Philosophical Studies* 161 (1): 17–26.

Engel, P. 2013b. Doxastic Correctness. *Proceedings of the Aristotelian Society* 87 (1): 199–216.

Engel, P. 2013c. In Defense of Normativism About the Aim of Belief. In *The Aim of Belief*, ed. T. Chan, 32–63. Oxford: Oxford University Press.

Engel, P. 2015. Epistemic Norms and the Limits of Epistemology. In *Hinge Epistemology*, ed. A. Coliva and D. Moyal-Sharrock, 15–175. Leuven: Brill.

Evans, J., and K. Frankish (eds.). 2012. *In Two Minds*. Oxford: Oxford University Press.
Findlay, J.N. 1954. The Justification of Attitudes. *Mind*, LXIII (250): 145–161.
Findlay, J.N. 1962. *Values and Intentions*. London: Allen and Unwin.
Gendler, T. 2008. Alief and Belief. *Journal of Philosophy* 105 (10): 634–663.
Gilbert, M. 2013. *Joint Commitment*. Oxford: Oxford University Press.
Glüer, K., and A. Wikforss. 2009. Against Content Normativity. *Mind* 118 (469): 31–70.
Hazlett, A. 2013. *A Luxury of the Understanding*. Oxford: Oxford University Press.
Hieronymi, P. 2008. Responsibility for Believing. *Synthese* 161 (3): 357–373.
Holton, R. 2014. Intention as a Model for Belief. In *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*, ed. Manuel Vargas and Gideon Yaffe, 12–37. Oxford: Oxford University Press.
Kahneman, D. 2012. *Thinking, Fast and Slow*. London: Penguin Books.
Kaplan, M. 1996. *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.
Kenny, A. 1963. *Action, Emotion and Will*. Oxford: Oxford University Press.
Koriat, A. 2005. Knowing and the Feeling of Knowing. In *Synergies*, ed. W. Østreng, 89–92. Oslo: Centre for Advanced Study.
Leitgeb, H. 2017. *The Stability Theory of Belief*. Oxford: Oxford University Press.
Lycan, W. 1985. Tacit Belief. In *Belief*, ed. R. Bogdan, 61–82. Oxford: Oxford University Press.
Maloney. 1990. It's Hard to Believe. *Mind and Language* 5: 122–148.
Mc Hugh. 2012. The Truth Norm of Belief. *Pacific Philosophical Quarterly* 93: 8–30.
Meinong, A.V. 1902. *Ueber Annahmen*. Leipzig: J. A. Barth. Reprinted in G*esamtausgabe*, ed. R. Haller and R. Kindinger. Graz: Akademische Druck- u. Verlagsanstalt. Geseimsaugabe, 1968–78, vol. IV, 385–489.
Mill, J.S. 1843 (1973, 1974). *A System of Logic; Collected Works of John Stuart Mill*, vols. 7 and 8, ed. J.M. Robson. Toronto and London: University of Toronto Press, Routledge.
Mulligan, K. 2007. Intentionality, Knowledge and Abstract Objects. *Disputatio* 2: 205–228.
Owens, D. 2003. Does Belief Have an Aim? *Philosophical Studies* 115: 283–305.
Papineau, D. 2013. There Are No Norms for Belief. In *The Aim of Belief*, ed. T. Chan, 64–79. Oxford: Oxford University Press.
Reisner, A. 2008. Weighing Pragmatic and Evidential Reasons for Belief. *Philosophical Studies* 138 (1): 17–27.
Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.
Scanlon, T. 1998. *What We Owe to Each Other*. Harvard: Harvard University Press.

Schwitzgebel, E. 2001. In Between Believing. *Philosophical Quarterly* 51: 76–82.
Shah, N. 2003. How Truth Governs Belief. *Philosophical Review* 112 (4): 447–482.
Sosa, E. 2015. *Judgment and Agency*. Oxford: Oxford University Press.
Sperber, D. 1985. Apparently Irrational Beliefs. In *On Anthropological Knowledge*, 35–63. Cambridge: Cambridge University Press.
Steglich-Petersen, A. 2010. The Truth Norm and Guidance: A Reply to Glüer and Wikforss. *Mind* 119 (475): 749–755.
Stich, S. 1978. Belief and Subdoxastic States. *Philosophy of Science* 45: 499–518.
Stout, G. 1896. *Analytic Psychology*, 2 vols. London: Sonnenschein.
Teroni, F. 2007. Emotions and Formal Objects. *Dialectica* 61 (3): 395–415.
van der Schaar, M. 2013. *George Stout and the Origins of Analytic Philosophy*. London: Palgrave.
Velleman, D. 2000. On the Aim of Belief. In *The Possibility of Practical Reason*, 244–281. Oxford: Oxford University Press.
Williams, B. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.
Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
Zimmerman, A. 2007. The Nature of Belief. *Journal of Consciousness Studies* 14 (11): 61–82.

PART III

Language, Contextualism and Naturalism

Naturalizing Picardi

Diego Marconi

Eva Picardi was not just a deep and perceptive expositor of Frege's and Dummett's philosophies of language, she was a philosopher of language in her own title. In this capacity she wrote a number of papers dealing with issues that I am also much concerned with, such as the public or social character of language, understanding, and knowledge of language (or, in my own terminology, semantic competence). In dealing with such issues she criticized Davidsonian semantic individualism, naturalism of the Chomskyan variety, deferentialism, and, more recently, some varieties of contextualism (such as Travis' and Recanati's views). Eva was a (moderate) *literalist*, as opposed to a radical contextualist:

> That a grasp of the meaning of a word that we have encountered on previous occasions may often be *insufficient* to determine how the word in question is meant …does not imply that there is no such thing as understanding the meaning of a sentence type outside a context of use. Admittedly, the understanding of a sentence type does not issue in the understanding of a context-invariant *proposition*, but it definitely narrows down the choice of eligible candidates. (Picardi 2010: 172)

D. Marconi (✉)
Department of Philosophy and Education Studies, University of Turin, Turin, Italy
e-mail: diego_marconi@fastwebnet.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_15

She was also a *normativist*: against Davidson, she did not believe that the rejection of (Burge's and Putnam's) deferentialism entailed "that appeal to the literal or standard meaning that words have in common language is spurious" (2006: 399). She insisted on the notion of *correctness*: implicit or explicit reliance on what we take to be the linguistic standard is constantly operative in linguistic practice (for example, in reporting malapropisms we usually correct them, i.e. we express what we take to be the intended meaning in the standard form, 2006: 399); and she feared that the crucial distinction between what seems correct and what is correct may be lost in semantic individualism (2006: 394). Finally, she was a *socialist*, i.e. she believed that any theory of language must account for "language as a social practice and as a cultural phenomenon" (1997: 121).

I happen to agree with Eva on all such views, and I share her philosophical targets. On the other hand, I am much more of a naturalist than she was. In what follows, I will try to reconcile what I regard as essential in Eva's views with the kind of moderate naturalism that I favour. I will also occasionally criticize some of her views, that I do not regard as essential.

I shall focus on two connected topics:

- First, what is the correct way of criticizing Chomsky from a normativist and socialist viewpoint;
- Secondly, what account of the social character of language must be opposed to semantic individualism.

I

Many of Eva's antinaturalistic remarks were made in connection with Chomsky's theory of language as presented in two well known and much discussed articles of the mid-Nineties, "Explaining language use" (Chomsky 1992) and "Language and nature" (Chomsky 1995; both later subsumed in Chomsky 2000). In those articles Chomsky had criticized a large part of the most influential philosophy of language of the previous two decades, including the work of Kripke, Putnam, Burge, Davidson, Dummett, and many others. Besides, he there presented his own view of the object and method of a scientific account of language (as opposed to philosophical speculations about something philosophers call 'language'; speculations that, besides being flawed in many ways, are anyway not concerned with a scientifically viable notion of language).

For Chomsky, the *object* of a scientific account of language is the procedure that generates structural descriptions, i.e. complexes of phonetic, semantic, and structural features; linguistic expressions may be identified with such structural descriptions. The generative procedure itself is called an I-language, where 'I' stands for internal, individual, and intensional (Chomsky 2000: 26). "Internal", to stress distinction from external phenomena such as inscriptions or patterns of sound waves, that are sometimes, though wrongly, identified with language; by contrast, an I-language is "a property of the brain" (Chomsky 2000: 27). "Individual", in that we are studying, and can only be studying the generative procedure of an individual speaker, something close to what we usually call an idiolect. "Intensional", meaning that an I-language is not individuated by the class of descriptions it generates (in principle, two different I-languages could generate the same class of expressions; thus an I-language is emphatically *not* a class of expressions); an I-language is individuated computationally, as a particular effective procedure. A systematic description (i.e. a theory) of an I-language is a *grammar* for that language (Chomsky 2000: 5). In words that are closer to those Chomsky himself more recently adopted, a grammar specifies sound-meaning pairings (as characteristic of a given idiolect), i.e. which sounds correspond to which meanings. Meanings should be understood mentalistically: they are mental entities constrained by the properties of a subsystem of the mind that Chomsky calls 'conceptual-intentional'. Other animal species are also endowed with analogous systems. The function that specifies the sound-meaning pairings is not the function that a speaker computes either in production or in comprehension (as the latter does not generate sound-meaning pairs but sounds *from* meanings, or vice versa); however, the former function (i.e. the grammar) determines the functions that speakers actually compute, in that it fixes the pairings that are each of them computed by a speaker (Matthews 2006: 208). The grammar determines the extensions of such functions, though not their intension, i.e. the algorithms that effect the computation: grammatical theory does not claim to determine *how* a speaker computes the sound that is paired with a meaning. In this sense, the theory of language does not tell us how we come to express a given thought by a certain sound pattern.

A speaker is said (or was still said in Chomsky 2000, e.g. 50–51) to have *knowledge* of her own I-language. Not, however, in the sense that "some sort of cognitive relation holds between Jones and his language, which is somehow 'external' to Jones" (Chomsky and Stemmer 1999:

397), nor in the sense that Jones has "a theory" of his language (ibid.). The knowledge a speaker has is said to consist in "mastery and internal representation of a specific I-language", where, however, 'representation' is not to be understood relationally, as 'representation of' (Chomsky 2000: 159–160). Perhaps the word 'representation' ought to be simply dropped: it does not appear to add much to the idea of *having* a language, in the sense in which one may be said to have sight, or even to have (rudimental) arithmetic, as properties of one's brain.[1] (Chomsky himself pointed out that English uses "knowledge of language where other…linguistic systems use such terms as 'have a language', 'speak a language'" (Chomsky and Stemmer 1999 cit.)). In a moment, as I discuss Eva's criticism I'll try to make the Chomskyan picture clearer.

Provocatively, for Chomsky language is not essentially for communication; even more provocatively, it did not evolve for communication. However, Chomsky does not deny that we do communicate by means of language. His account of communication is straightforward:

> It may be that when he listens to Mary speak, Peter proceeds by assuming that she is identical to him, modulo M, some array of modifications that he must work out. Sometimes the task is easy, sometimes hard, sometimes hopeless. To work out M, Peter will use any artifice available to him, though much of the process is doubtless automatic and unreflective… Insofar as Peter succeeds in these tasks, he understands what Mary says as being what he means by his comparable expression. The only (virtually) "shared structure" among humans generally is the initial state of the language faculty." (Chomsky 2000: 30)

Thus to understand some linguistic performance based on a (more or less) different I-language we work out certain adjustments—one presumes, with respect to the default functioning of our computational device that associates meanings to sounds—and for that we may "use any artifice available" (i.e. no part of our cognitive system is in principle excluded), though "much of the process is automatic"—this means, I believe, that many modifications do not surface to consciousness and/or they are taken care of by the computational device itself.

Most of Eva's criticisms of Chomsky can be brought back to either of two objections. First, Eva believes that Chomsky's notion of "unconscious", or "tacit" knowledge of language is as flawed as Dummett's notion of "implicit knowledge" of language (a notion Dummett himself

had eventually given up), and for the same reasons. Secondly, she claims that Chomsky's picture does not account for a number of important features of communication.

Concerning implicit knowledge, Dummett had pointed out (in his "Preface" to *The Seas of Language*, Dummett 1993) that, first, if knowledge of language is implicit then it is not clear how the subject can apply it in language understanding (or production, for that matter), as such knowledge is supposedly not cognitively available to her. Secondly, implicit knowledge is irrelevant to (Dummett says "incompatible with", which may be unnecessarily strong) the characterization of language as a practical ability. We can learn how to do something, says Dummett—i.e. we can acquire a practical ability—without thereby acquiring theoretical knowledge, implicit or otherwise, of how we go about doing it. Hence, language as a practical ability need not be based on implicit theoretical knowledge, nor does the acquisition of language, as a practical ability, generate theoretical knowledge of language. Therefore, the notion of implicit knowledge plays no explanatory role whatsoever.

Eva appears to think that such criticism applies to Chomsky's notion of tacit knowledge of an I-language. However, Dummett's second argument essentially depends on his definition of implicit knowledge as knowledge whose content a subject cannot formulate, though he can recognize as correct a formulation that is offered to him (1993: xi). It is because implicit knowledge can be so recognized that Dummett can show that acquiring a practical ability does not involve the acquisition of implicit theoretical knowledge: for the subject, even if he has acquired the practical ability, more often than not does *not* recognize a formulation of "how he goes about doing it" as correct. Hence, we are licensed to conclude that he does not possess such implicit knowledge. The argument, however, does not apply to tacit knowledge in Chomsky's sense: as far as I know, Chomsky never claimed, and it would have been quite implausible for him to claim that a subject would recognize a formulation of a grammatical principle (say) as correct when offered one (see Chomsky 1986: 265).

Dummett's first argument, on the other hand, is inherently dubious. Granted that implicit knowledge must be available *to the cognitive system* (or to some of its components) if it is to play a role in linguistic performance, it is not clear that it must be available *to consciousness*, i.e. that the subject must be able to become aware of the content of such knowledge, as Dummett appears to be assuming. In any case, Chomsky does not see

it that way. One way he put it is this: the application of a subject's knowledge of her grammar stems from quasi-deductive processes from the principles of the grammar; though the conclusions of such "deductions" may or may not be conscious (Chomsky believes that in some cases they are, while in other cases they definitely aren't), the quasi-deductive process itself is not transparent to consciousness (1986: 270). Dummett and Picardi would surely be right in pointing out that we still have no proper *theory* of how tacit knowledge of language is put to use in language production and language understanding; however, this is not to say that the very fact that knowledge of an I-language is tacit is incompatible with such knowledge being applied.

None of this amounts to endorsing Chomsky's views of what it is to know one's language, or to suggesting that such views are entirely clear. For example, like other commentators I am quite unhappy with Chomsky's persistent use of the word 'knowledge' to designate a subject's relation with his I-language.[2] My point here is just that Dummett's criticism of his own former notion of implicit knowledge of language, aside from being partly unconvincing, does not apply to Chomsky's tacit knowledge.

What about Eva's criticism of Chomsky's picture of communication? She makes three points (1997: 114):

1. In Chomsky's account, no role is played by "the way in which language is interwoven with all other *activities* which we perform"; but it is hard to believe that the connections between linguistic utterances and the rest of our life play no role in communication.
2. Chomsky says that we understand other people by introducing suitable "modifications" with respect to our own language, but our I-language is said to be "largely inaccessible" to consciousness; how, then, can we modify it?
3. How, on Chomsky's picture, do we manage to understand "turns of phrases which seem unintelligible if taken literally", or phrases which, though intelligible, do not chime with what the speaker later says or does, with the way he "acts on" his own or other people's utterances?

To evaluate such criticisms, let me go back to the Chomskyan picture. On Chomsky's view, we are endowed with some mechanism that allows us to express thoughts by means of sounds—a computational device

that associates meanings with phonetic forms. The mechanism works both ways: it also allows us to associate a meaning to a phonetic form (a string of sounds). We have a theory of such associations, as we have a computational theory (in Marr's sense)—a grammar—that generates the sound-meaning pairings. We do not, however, have a theory of how the individual speaker/hearer goes about associating a sound to a meaning, or a meaning to a sound: we don't know the algorithms she implements. Now, the issue arises of how we go about understanding an interlocutor whose I-language is not identical with ours, i.e. does not associate exactly the same meanings with the same phonetic forms. That we do understand these interlocutors shows that we can make adjustments for small variations in such associations. Some adjustments can be taken care of by the computational device itself (for example, the device can be pretty robust as to variations in pronunciation and prosody—though mine is not, I admit). In such cases, *we* do not make the adjustments; they are part of the mode of operation of the computational device. So there is no issue of such modifications being or not being possible *for us*, as an I-language is impenetrable to consciousness: the adjustments are as impenetrable as any other aspect of the device's mode of operation.

But then, there are cases in which we explicitly and consciously conjecture what the speaker may have meant by the words he uttered, as his words are "unintelligible if taken literally", as Eva says; or again, because his words, "taken literally", do not fit the rest of his behaviour, both linguistic and non-linguistic. Thus, for example, we reflectively conjecture that, say, Mrs Malaprop may have meant *epithet* by 'epitaph'. In such cases, says Chomsky, we use "any artifice available" (Chomsky 2000: 30): i.e. we make use of our cognitive resources to find a suitable *replacement* for the words that were in fact uttered; a replacement that, once processed by our computational device, yields a meaning that is no longer perceived as problematic in view of its incoherence with the rest of the speaker's behaviour or with general assumptions of rationality. Again, Eva's issue of how we can get to modify a computational device that is inaccessible to us does not arise. The device's mode of operation *is not modified at all*, in such cases: simply, its input is replaced by a different input yielding more satisfactory output. These are cases in which our general ability to cope with unexpected facts takes over; they have little to do with language understanding proper. We look for, and often find *reasons* to replace a certain sequence of sounds with a different sequence, which is then fed the computational device.

So I believe that answers can be provided, from Chomsky's viewpoint, to Eva's criticisms (2) and (3). Concerning (1), i.e. the objection that the Chomskyan picture makes no room for the obvious connections between language use and other human activities, it could be retorted that accounting for such connections is no business of a theory that explicitly downplays the communicative use of language (as I recalled above).[3] For the connections at issue are likely to involve, not language per se, but its communicative uses. Still, I also find something unconvincing in the picture: something that is related to another theme of Eva's polemic against semantic individualism, namely the issue of normativity and the social character of language.

Robert Matthews (2006) pointed out that, in coherence with the Chomskyan account, there is another sense in which a speaker may be said to know his own I-language: speakers can issue authoritative judgements as to which sound is paired with which meaning. Although speakers do not have conscious and explicit knowledge of the grammar, i.e. of the computational theory that determines sound-meaning pairings, they know how to effect such pairings and, as a consequence, they can issue judgments to the effect that a sound is paired with a certain meaning (they have access to the output of the computational device). We do have access to the output of mechanisms whose functioning eludes consciousness: for example, sometimes our body tells us, though indirectly and inexactly, that we have high pressure, or that our digestion went wrong somehow. But in the case of our idiolect, we are told with great exactness what a certain sound means or how a certain meaning is to be voiced; and we are so told "for uncountably many possible utterances" in our idiolect (Matthews 2006: 216). It's rather like the case of vision: "Our linguistic competence grounds our authority [in the case of language] in roughly the way that our visual competence grounds our judgments about what we see" (Matthews 2006: 217).

Here, the linguistic computational device is assimilated to the visual system in that, like the visual system,

- it is an automatic, subpersonal mechanism whose operations are opaque to consciousness,
- its output is available to consciousness,
- indeed, the subject has first person authority on it. One cannot be wrong on what a sound means in one's idiolect, any more than one can be wrong about what it appears to one that one sees.

There are, however, two important differences between vision and the linguistic device. First of all, much of language is (not just evolved, but) *learned*. In Paul Bloom's words, no one is born associating the sound |rabbit| with a certain meaning (Bloom 2002: 15). Whereas in the case of vision, we do not learn to associate a certain pattern of stimulation on the retina with the visual image of this room.

Secondly, though we may mistrust our sight in special circumstances, we cannot, under any circumstance, *correct* our vision mechanism voluntarily (though we may undergo surgery to have it corrected). Whereas in the case of language, we can and do modify our pairing device if we are told that something is wrong with it. For example if I am told that, contrary to how things go in my idiolect, the sound |presently| ought not to be associated with the meaning of 'at present' I do not have to undergo surgery to modify my linguistic device. If I choose to regard my corrector as authoritative I simply go ahead and modify my idiolect: from now on, the sound |presently| will be paired with the meaning of 'in a short while'.

Both differences are obviously connected with each other. It is inherent to learning proper (and learning lexical meanings is learning proper) that one can both learn from an unreliable source and *mis*learn, i.e. learn something different from what one is intended to be taught. Hence there is room for correction, from better sources or better understanding of the teaching source.

Both of these features (which characterize language as opposed to vision) are sort of peculiar for an automatic, brain-implemented mechanism to have. But, leaving that aside, what this shows is, first of all, that the linguistic computational device is subject to modifying external influence in a way in which the system of vision is not; moreover, such modifying influence is not limited to the early parameter-fixing stage, but goes on throughout an individual's life. An account of I-language should explain such permeability to change. *Why* do we accept to be corrected? Even before that, why do we regard other people's normative indications as relevant to our idiolect? Why don't we just keep talking the way we used to talk—the way our I-language dictates—and let Davidson's "wit, luck, and wisdom" (1986: 173) take care of other people's understanding us? Eva is quite right in pointing out that Chomsky's picture does not provide an answer to these questions.

II

Now, there is a popular answer to the above questions (popular among Davidsonians, at any rate): we change our ways "to make communication smoother". If we speak like other people, i.e. if our I-language is more similar to other people's I-languages, a smaller effort is required of their cognitive systems to understand what we say; this makes communicative interactions easier and faster, which is a good thing (see e.g. Bilgrami 1992: 111–112). There are several objections to this answer.[4] First of all, at least *prima facie* we do not at all modify our way of speaking *in order to speak like other people*; i.e. the norm we feel we obey is not "Speak like other people!", but rather "Speak as you *ought to* speak!"—i.e. conform to standard, no matter what other people may do. Secondly, and most importantly, this account presupposes that there is an alternative: if we gave up politeness and the reasonable wish to make communication smoother, we could refrain from changing our ways—in general, we could perfectly well "talk the way we talk", i.e. in any way whatsoever, and be understood nevertheless. After all, even "most of "The Jabberwock" is intelligible on first hearing" (Davidson 1986: 158).[5]

Among many surprising claims that Davidson put forth in his philosophical career this is, in my opinion, one of the most surprising. And that not so much because it is obviously false: many people feel they cannot make any sense of much of "Jabberwocky" even on third hearing, or reading. The point is, rather, that *there is no sense to be made*: the pseudo-words that compose the poem were not intended by Carroll to have any meaning. Lewis Carroll may well have chosen his pseudo-words on the basis of phonetic affinities with genuine words, and he may have intended stylistic effects similar to those one would get using genuine words; some of the pseudo-words may have carried semantic associations for him, and they may carry associations in the reader's mind, which may or may not coincide with Carroll's associations. But none of this entails (1) that Lewis Carroll had a definite communicational intention in mind, like Mrs Malaprop, or (2) that the pseudo-words in "Jabberwocky" uniquely, or even with probability evoke familiar, genuine words. "Jabberwocky" was *not intended to be* "intelligible", and should not be, if one is to respect Carroll's real communicative intention.[6] Of course, one can treat grooves and engravings on an ancient temple's wall as characters of a forgotten alphabet and attribute meaning to them: this is not to say that such engravings are intelligible.

"Jabberwocky"'s case is not like the cases we were talking about: we were talking about people who have genuine contents to communicate and the intention to communicate them. Is it true that they would manage to communicate them no matter how they chose to express them? No. If I chose to write the rest of this paper in a mixture of Urdu and Swahili, most of my readers wouldn't understand me in spite of any amount of contextual evidence and familiarity with my topic, not to mention wit and wisdom; readers would need an incredible amount of luck (Davidson's third factor) just to figure out whether I am still talking philosophy or I turned to philately instead. So, we have no alternative to speaking like other people (*to a large extent* like other people) if we want to make communication not smooth, but possible.

On the other hand, we do get away with a certain amount of differences. An interesting question is, why do we care about such differences *as well*? Why do we accept to stand corrected concerning the meaning of 'profligate' or 'pellucid' (or 'epitaph')?

Let's put the issue in slightly different terms. Is it invariably the case that communication is about recovering a speaker's or writer's communicative intentions? No, Eva says:

> [T]he issue of understanding what a sentence says must be kept separate from that of understanding the content of the specific belief which a speaker may want to convey to a specific hearer by uttering it under certain circumstances. (Picardi 1997: 119–120)[7]

Take the case of legally binding (or even legally relevant) written texts. With such texts, we do not just care about what the person who wrote the text meant, or intended, or had in mind (though that may be relevant as well): what we mostly care about is what the words she used mean in English. This is why linguists, and even philosophers of language are sometimes summoned in courts, to explain not what a speaker or writer had in mind—they obviously wouldn't know—but what *the text*, or *the uttered words*, *mean*.

Now, perhaps fortunately, not all we say or write has legal import. However, as producers of texts and spoken words we all risk being interpreted by interlocutors who do not much care, and are not supposed to care about what we may have meant by them; what they care about is what we actually *said*. Now, we could dismiss such an uncooperative attitude, were it not for the fact that, in such cases, it is generally agreed that

we should *stand by* our words, or that we are *responsible* for what we said (not for what we meant); i.e., it is generally agreed that the workings of our I-language are irrelevant for the practical consequences of our speech acts, including legal responsibility and moral evaluation.

This being so, we are justifiably concerned about guaranteeing, to the greater possible extent, that our linguistic productions will be interpreted as we want them to be interpreted. But how can we force other people to interpret our words in one particular way? I-languages differ (though perhaps not widely among our usual interlocutors, still widely enough to entail catastrophic consequences). Nothing can protect us against the vagaries of interpretation. But, luckily, no such protection is necessary. For if the interpretation we want our words to receive is the *standard* interpretation, then, even if that is not the interpretation they will de facto be given by the interpreter, we will not be held accountable for such a de facto interpretation. We will be taken to be responsible for the standard interpretation of our words, not for any other interpretation that may be placed upon them by this or that interpreter. Alternative interpretations *will count as misunderstandings*: we will not be held accountable for the meanings they assign to our words. Hence, the only way we can guarantee that the interpretation of our spoken or written words will not get us in trouble is by having the standard interpretation be the one we want our words to receive.

But then, the standard interpretation must suit us: we must be prepared to stand by our words in the standard interpretation. So our words, in the standard interpretation, should express the contents we want to convey: not necessarily our real opinions or our deepest beliefs, of course, but anyway the thoughts we are willing to be held accountable for. I.e., it is best for us to have the standard interpretation of our words to coincide with the thoughts we are willing to have ascribed to us and, consequently, held accountable for. Briefly put, we have good reasons to aim at *speaking according to standard*. We want it to be so that the standard interpretation of the words we utter coincides with the thoughts we want to convey (as they are the thoughts we are prepared to stand by and be judged for). If we expressed the thoughts we want to convey in a non-standard linguistic form, we would risk being judged for entertaining thoughts that differ from those we are prepared to stand by. This is why, no matter how we as a matter of fact speak or write, we usually *believe* to be speaking or writing according to standard: if we didn't think so we would speak and write differently, for most of the time[8] we *aim* at speaking and writing according to standard, for very good reasons.

Of course, nothing of this entails that we *do* speak according to the standard (in fact, we often don't). However, it does explain a few of things. First of all, it explains why we accept corrections concerning the meaning of 'profligate' or 'pellucid'. There are circumstances in which it may be crucial for us to speak according to standard; but then, one never knows: maybe in some such circumstances even the word 'profligate' will make a difference. So it may be useful to know the standard in some detail, and be prepared to conform even in such detail. Secondly, it explains why *there is* a standard. There is a standard so that people will not be held responsible for whatever meanings an interpreter attaches to their words, particularly in cases in which interpreters are neither willing nor obliged to care about what a speaker may have *meant* by her words—they are not exercising their "wit, wisdom, and luck", nor are they supposed to. Thanks to the standard, people can rest assured that their words will not be subject to the vagaries of interpretation, as far as their responsibilities are concerned. The existence of a standard is not required to distinguish between understanding and misunderstanding: intention and interpretation suffice, as a misunderstanding can be defined as an interpretation that betrays the speaker's intentions. However, reference to a standard is required for an interpretation to count as a misunderstanding *independently of a speaker's intentions*, hence even where interpreters are not under an obligation to worry about them.

Thirdly, it explains what the standard is. The standard is a sound-meaning pairing which may or may not coincide with some individual speaker's pairing—that does not matter. Knowledge of it is conveyed by a number of publicly available devices (e.g. dictionaries) that rely for their effectiveness on individual speakers' I-languages. Such devices can be compared to eyeglasses. The big difference, of course, is that in the case of language, modification of the input-output pairing depends on individual deliberation, whereas in the case of vision it is governed by physical laws and biological necessity: deliberation is confined to input modification.

III

Essentially, I have been re-telling a story that has been told before, by both Eva and Michael Dummett. Dummett pointed out that a speaker's commitments and the ensuing responsibilities are determined by the standard interpretation of her words:

> When an utterance is made, what the speaker *says* depends upon the meanings of his words in the common language; but, if he thereby expresses a belief, the content of that belief depends on his personal understanding of those words, and thus on his idiolect. (Dummett 1991: 88)

Eva insisted that a hearer is not necessarily pursuing understanding of what a speaker *means* on a given occasion:

> [T]he issue of understanding what a sentence says must be kept separate from that of understanding the content of the specific belief which a speaker may want to convey to a specific hearer by uttering it under certain circumstances. (Picardi 1997: 119–120)[9]

I have been trying to show how both facts taken together explain why are we sensitive to the standard to the point of pursuing conformity even in details, granted that, in many cases, our interlocutors are likely to get what we mean even when we do not conform.

This account highlights the limits of semantic individualism. Though the individualist does not dispute the phenomena of semantic normativity, he cannot find any motivation for them other than the desire to make communication more efficient. This is because, in his view, the whole point of language is communication and the whole point of communication is conveying/understanding what a speaker means. The semantic individualist does not see that, in a number of cases, what matters is not so much what a speaker means as what her words mean. In such cases, conformity to the standard is not optional: it is the only way we can ensure that we will not be held responsible for contents we are not prepared to stand by. This is why we pursue conformity, though we may not always achieve it.

We saw that an I-language is subject to deliberate modifications in a way in which, for example, sight is not. At least some such modifications are motivated by social needs, and their content depends on a social institution, namely the semantic standard. Does this entail that an I-language is not a natural object but some sort of social construction, an artificial object like writing or the digital computer? Eva would probably have answered that, no matter what an I-language might be (if there is such a thing), *language* is not a natural object. I don't see that any such straightforward conclusion follows. If Chomsky is right about grammar and the way it determines an individual I-language (and nothing I said

so far contradicts him on this), then much of an I-language is anyway fixed by natural constraints; such constraints extend to what we call the lexicon, by limiting possible sounds, possible meanings, and even possible sound-meaning pairings at the lexical level.[10] Within such limits, there is no doubt considerable room for different choices in the pairing of sounds and meanings. However, we do not stop regarding walking as a natural, biological process once we realize that its direction and speed are not biologically determined. Similarly, an account of semantic normativity such as the one I have proposed is compatible with regarding an I-language as a natural function.

Obviously, the Chomskyan account of language is predicated on a conception of meanings as mental entities, constrained by a subsystem of the mind that is called the "conceptual-intentional system". This is at odds with views on which meaning is use, and lexical meanings are either *patterns* of use of certain sounds, as in Paul Horwich's theory of meaning (Horwich 1998, 2005), or *norms* for the use of certain sounds (as in many accounts inspired by Wittgenstein, e.g. Glock 2010). Or so it seems. In fact, I wonder whether even conceptions of meaning as use could not be reconciled with an essentially naturalistic account of linguistic production and understanding. This, however, is a matter for another paper.

Notes

1. "Knowledge of language - says Chomsky - [is] the internal representation of [the generative procedure] in the brain" (2000: 50).
2. For example: "After the development of [the principles and parameters approach] it really became transparent that grammars and/or languages could no longer be sensibly thought of as independent objects of knowledge. …Grammars, and so languages, ceased to be understandable as things which speakers/hearers know; they are simply states of the speaker/hearer" (Collins 2004: 512).
3. The view that human language is not primarily for communication is famously reiterated in Hauser et al. (2002).
4. Some of them I already put forth in Marconi (1997, chapter 4).
5. "The Jabberwock" (or rather, "Jabberwocky") is a nonsensical, pseudo-Medieval poem created by Lewis Carroll in *Through the Looking-Glass*. It begins: "'Twas brillig' and the slithy toves/Did gyre and gimble in the wabe".

6. Both Humpty Dumpty (in the novel) and Lewis Carroll (in letters and other literary works) provide "explanations"—often diverging from one another—of several words in the poem. These are not to be read as genuine semantic interpretations exhibiting intended meanings, but as more literary invention. See Gardner (2001: 157–164), Heath (1974: 139).
7. See also the discussion of Grice in Picardi (1999, particularly pp. 98–99).
8. Jokes, word play, and deliberate pretense are obvious exceptions.
9. See also the discussion of Grice in Picardi (1999, particularly pp. 98–99).
10. For example, it has been argued that no transitive verb in English could express the converse relation of *break*: *The desk blikked Mary* could not mean that the desk was broken by Mary. See Johnson (2004) and the literature he quotes.

References

Bilgrami, A. 1992. *Meaning and Belief*. Oxford: Blackwell.
Bloom, P. 2002. *How Children Learn the Meanings of Words*. Cambridge: MIT Press.
Chomsky, N. 1975. *Reflections on Language*. New York: Pantheon.
Chomsky, N. 1986. *Knowledge of Language*. New York: Praeger.
Chomsky, N. 1992. Explaining Language Use. *Philosophical Topics* 20: 205–231.
Chomsky, N. 1995. Language and Nature. *Mind* 104: 1–61.
Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
Chomsky, N., and B. Stemmer. 1999. An On-Line Interview with Noam Chomsky: On the Nature of Pragmatics and Related Issues. *Brain and Language* 68: 393–401.
Collins, J. 2004. Faculty Disputes. *Mind and Language* 19: 503–533.
Davidson, D. 1986. A Nice Derangement of Epitaphs. In *Philosophical Grounds of Rationality*, ed. R. Grandy and R. Warner, 157–174. Oxford: Oxford University Press.
Dummett, M. 1991. *The Logical Basis of Metaphysics*. London: Duckworth.
Dummett, M. 1993. *The Seas of Language*. Oxford: Clarendon Press.
Gardner, M. 2001. *The Annotated Alice*. London: Penguin Books.
Glock, H.-J. 2010. Does Language Require Conventions? In *Wittgenstein: Mind, Meaning and Metaphilosophy*, ed. P. Frascolla, D. Marconi, and A. Voltolini, 85–112. London: Palgrave Macmillan.
Hauser, M.D., N. Chomsky, and W.T. Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298: 1569–1579.
Heath, P. 1974. *The Philosopher's Alice*. New York: St. Martin's Press.
Horwich, P. 1998. *Meaning*. Oxford: Clarendon Press.
Horwich, P. 2005. *Reflections on Meaning*. Oxford: Clarendon Press.

Johnson, K. 2004. Impossible Words. *Mind and Language* 19: 334–358.
Marconi, D. 1997. *Lexical Competence*. Cambridge, MA: MIT Press.
Matthews, R. 2006. Knowledge of Language and Linguistic Competence. In *Philosophy of Language*, ed. E. Sosa, E. Villanueva, 200–220. Boston & Oxford: Blackwell.
Picardi, E. 1997. Is Language a Natural Object? In *Thought and Ontology*, ed. M. Sainsbury, 107–123. Milano: Angeli.
Picardi, E. 1999. *Le teorie del significato*. Roma: Laterza.
Picardi, E. 2006. Individualismo semantico e significato letterale. In Le ragioni del conoscere e dell'agire, ed. R.M. Calcaterra (a cura di), 392–402. Milano: Angeli.
Picardi, E. 2010. Wittgenstein and Frege on Proper Names and the Context Principle. In *Wittgenstein: Mind, Meaning and Metaphilosophy*, ed. P. Frascolla, D. Marconi, and A. Voltolini, 166–187. London: Palgrave Macmillan.

Practical Knowledge and Linguistic Competence

Annalisa Coliva

1 Introduction

One of Eva's classes I took in the early 1990s was on the topic of rule following. In those very years, the late Roberto Dionigi was teaching on Ludwig Wittgenstein. Eva and Roberto were united by deep friendship, great professional respect, and by the propensity to take the "later" Wittgenstein very seriously. Those classes influenced me deeply, even when I turned to epistemology. The present chapter attempts a synthesis between those apparently distant areas of philosophy—epistemology and the philosophy of language—in a broadly Wittgensteinian perspective. I will focus on the topic of our knowledge especially of the syntactic

My deepest thanks go to Danièle Moyal-Sharrock and Aaron James for comments and criticisms on the penultimate draft of this chapter. The first version of the chapter was originally presented in French at Collège de France. I wish to thank Claudine Tiercelin for inviting me to deliver a talk in the context of her seminar. Thanks are also due to Andrea Sereni and to people in attendance at a talk delivered at Istituto Universitario di Studi Superiori in Pavia.

A. Coliva (✉)
University of California Irvine, Irvine, CA, USA
e-mail: a.coliva@uci.edu

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_16

rules of a language and I will address this problem by considering Noam Chomsky's position, according to which, such knowledge is propositional and innate—or, to use Michael Dummett's (1981) terminology, "unconscious". Eva herself (Picardi 1997, 2001) addressed Chomsky's views on language, and developed interesting comparisons between Chomsky, Davidson and Dummett, favoring Dummett's position with respect to language.[1] Here, I will focus mostly on Chomsky's view that linguistic knowledge is propositional. This conception is explicitly opposed to the one, developed by Gilbert Ryle (1949, 1971) and Ludwig Wittgenstein (1953, 1969), according to which the rules that govern our linguistic behavior are acquired through learning and their knowledge constitutes an example of practical knowledge, or know how. Although for Wittgenstein and Ryle these rules can be made explicit, at least in some cases, in order to be a competent speaker one needs to know them practically. For, no matter whether one knows them propositionally, what counts is to be able to do what these rules prescribe. More specifically, it consists in doing what the rule *makes explicit*, not in having the rule in mind and in being guided by it as one speaks. Nor does it consist in having it stored somewhere in one's mind or, even less, in having it codified somewhere in a region of one's brain. This kind of know how, then, pre-exists, and is largely independent of, the ability to formulate the rule. But if the rule guides us neither explicitly, nor unconsciously, the very idea that it has to be there—in our minds, or even in our brains, either at the personal or at the subpersonal level—evaporates. Thus, the Wittgenstein-Ryle view calls for the demise of the idea that linguistic competence is *guided* by those rules. Still, speaking a language is a rule-governed practice in the following sense. Our linguistic practice is regular. That is to say, it is not random, it is socially shared and it is repeated over time. Hence, we can elicit a rule from it. The eventual formulation of the rule allows us to best explain and make sense of what is going on in our practice. Indeed, an explicit formulation of the rule can be used, at times, to have the rule "enforced", when someone breaks it. Yet, the rule does not pre-exist the linguistic practice and it is not what guides speakers when they behave in conformity with the rest of their linguistic community. Indeed rules can change if the practice changes. Yet, it is only if several people, over a certain amount of time, linguistically behave a certain way that a new linguistic rule is initiated. Clearly it thus makes little sense to think that linguistic competence

consists in having knowledge—let it be innate, implicit or unconscious knowledge—of a rule.

The general epistemological issue of the nature of practical knowledge and know how has recently come to the fore in the work of Jason Stanley and Timothy Williamson (2001). They have maintained that practical and propositional knowledge are not categorically distinct and that the former is a subspecies of the latter. The debate initiated by their provocative chapter has sometimes veered into elusiveness and it has suggested several compatibilist positions. In the rest of this chapter, I will side with Wittgenstein and Ryle in claiming that, if we take the idea of practical knowledge seriously, we have to acknowledge that it is irreducible to propositional knowledge, both in the case of linguistic competence and in general.

The structure of the chapter is as follows. I begin by introducing the distinction between practical and propositional (or theoretical) knowledge. Then I consider some objections to it, with special attention to the ones put forward by Stanley and Williamson. Afterwards, I briefly summarize Chomsky's position on linguistic competence—a position that, as mentioned, is based on the idea that linguistic competence consists in propositional knowledge. Finally, I criticize both the theoretical and the empirical arguments Chomsky puts forward in favor of his view and present some observations in favor of the idea that linguistic competence is ultimately practical.

2 Practical and Propositional Knowledge

The distinction between practical and propositional knowledge is often illustrated by means of examples such as "John knows that Paris is the capital of France" and "John knows how to ride a bike". In the former case, we can say that John knows that Paris is the capital of France because, roughly, (i) he believes the proposition that Paris is the capital of France, (ii) that proposition is true, and (iii) he has a justification (usually of a testimonial nature) to believe that Paris is the capital of France. In the latter case, in contrast, the fact that John knows how to ride a bike does not depend on his believing certain propositions (those that jointly describe everything that is necessary and sufficient to do in order to ride a bike) and on his having a justification for each of them. Rather, it depends on his being able to do what one is required to do in

order to ride a bike: he can get on a bike and balance, he can steer the handlebars, use the breaks, etc.

To know how to ride a bike, therefore, consists in having a series of abilities. Sometimes, however, these abilities cannot be manifested for a variety of reasons. Yet, they would still exist. For example, John can break his leg and be no longer able to ride a bike. This does not necessarily mean that he has lost the ability (or the capacity) to do so. Rather, for contingent reasons he cannot manifest his ability. However, if those impeding factors can be removed and if, once removed, John can still ride a bike (perhaps after a little bit of exercise and rehabilitation), he still counts as having the ability.

An account of practical knowledge in terms of abilities (or capacities) is sometimes enriched by the observation that we manifest the ability to do *x* not just by doing *x* but also by doing things which are connected to it. In particular, being able to explain how to do *x* is taken to be sufficient for having the ability.[2] Such a dispositional analysis of practical knowledge is sometimes considered superior to that in terms of abilities because it would allow us to say that we have actually lost the ability when we are not in a position to manifest it, without entailing that we no longer know how to do that thing. For example, John would still count as someone who knows how to ride a bike, even though, given his broken leg, he has no longer the ability to ride the bike, for he would still be able to explain how to do it.

Yet, if John does not have the ability to ride a bike because of his broken leg, it is not clear why we should say that he still knows how to ride a bike. If we still want to say so, I surmise, it is because we think that he retains the ability and that, once recovered from his injury, he will be able to manifest it again. Hence, we can retain the analysis of practical knowledge in terms of abilities, and say that if John has a broken leg, he still has the ability to ride a bike—he still knows how to do it—provided the impeding factors are passing ones and that, once gone, he can still ride the bike.

Clearly, the mere disposition to explain how to ride a bike (or how to do anything else) is neither sufficient nor necessary in order to know how to do that. Consider those experts in sports or music, who know everything, *theoretically*, about how to smash, in tennis, say, or about how to play an instrument, but who cannot actually do any of these things. The kind of know how we are after is not theoretical knowledge of how to do something. Rather, it is that kind of knowledge one

possesses just in case one is practically able to do that thing, even if one were momentarily prevented from doing it, due to contingent impeding factors.[3] If, in contrast, those inhibiting factors persisted, they would actually annihilate our practical knowledge, or know how, while they could leave our theoretical knowledge of how to do that given thing intact. A former tennis player can be no longer able to play tennis due to age limitations. Yet, she can still possess theoretical knowledge of how to do or not to do certain movements, or of how to tactically play a match, etc. and she may pass on this knowledge to other players. Even so, none of this turns her knowledge into a practical kind of knowledge. Rather, it remains a case of theoretical knowledge, acquired by reflection on a practice, of how to do all those things.

Consider, furthermore, how the relevant practical abilities are normally acquired. Very often, we need to teach the body, or a part of it, how to make certain movements. This training usually consists of preparatory and intermediate exercises, which eventually put us in a position to do what we are supposed to do. They also very often involve endless repetitions, so that the movements become "automatic". The training may be accompanied by some explanation of how to do certain things, but it need not and clearly it does not consist in acquiring an explanatory ability.

Hence, I submit that overlooking the distinction between practical and theoretical knowledge of how to do something has led several theorists to think that practical knowledge could consist in knowledge of a series of propositions. Ryle considers this a form of "intellectualism". In its simplest form, intellectualism has it that knowing how to ride a bike, how to speak a language or how to play tennis consist in knowing a series of propositions (perhaps in some particular way).

If practical knowledge is equivalent to knowing a series of propositions, it might seem that subjects should be able to produce or assent to them. Thus, if knowing how to ride a bike is equivalent to knowing propositions such as "I have to get on the bike", "I have to balance myself", "I have to steer the handlebars" and "I have to use the breaks to stop", etc., a way of testing this thesis would be to check whether those subjects who can actually ride a bike would be able to produce such a list of propositions, or at least assent to them. That does not seem to be the case, though. People who can ride a bike may not be able to produce (or just assent to) this list of propositions. Besides, they might be able to ride a bike long before having the conceptual sophistication

needed even to grasp these propositions. Children may be able to ride a bike at age three and yet have none or very few of the concepts they would need to grasp those propositions. Conversely, many people who do not know how to ride a bike would be able to produce the above-mentioned list of propositions. What people who know how to ride a bike can do and those who do not know how to do it cannot do is actually to perform the actions described in that list. Moreover, people tend to give the wrong answer to questions regarding how they do things they are actually perfectly able to do.[4]

A natural way for intellectualism to resist this argument is to reject the premise that any propositional knowledge is either explicit, or such that one could make it explicit if need be. According to Jerry Fodor, who supports this strategy, "certain of the anti-intellectualist arguments fail to go through because they confuse knowing that with being able to explain how" (1968, p. 634, quoted in Fantl 2012). Hence, according to him, practical knowledge could be an instance of *tacit* propositional knowledge, which could not be made explicit.[5]

In response, it has to be acknowledged that the empirical data at our disposal do not simply show that when people know how to do something, like riding a bike, they are unable to explain how they do it. Rather, they show that people actually *deny* that they do those things in the way our best explanations tell us they actually do them.[6] This seems *prima facie* decisive evidence against intellectualism. However, one might want to insist—and Fodor and Chomsky seem to agree with that—that tacit or "unconscious" knowledge is compatible with the fact that those who possess it deny its content.

The issue is clearly elusive. For intellectualists will insist that a subject possesses tacit knowledge of a proposition even when she is disposed to denying its content. Anti-intellectualists, in contrast, will insist that it looks like a leap of faith to hold that a subject knows that P, however tacit her knowledge might be, if she denies that P. They will maintain, instead, that that subject knows practically how to do something, while not knowing the proposition that P.

The most powerful argument put forward in favor of intellectualism is due to Stanley and Williamson (2001). What attracts their attention is the symmetry between practical knowledge ascriptions and attributions of other kinds of knowledge regarding subjects, places, times, reasons, etc. That is, the kind of knowledge-ascriptions that concern *wh-questions*. For instance, just as we say

(1) John knows how to ride a bike

we do say

(2) John knows where the bike is;
(3) John knows who was the last person to ride the bike;
(4) John knows why they stole the bike, etc.

An approach to relative clauses, which treats all of them on a par, has a pleasing theoretical simplicity. Furthermore, the relevant literature in linguistics seems to converge on the idea that these constructions should be treated in terms of propositional knowledge. Stanley and Williamson maintain that these sentences should be interpreted as follows.

(2*) John knows, of a place *l*, that *l* is the place where the bike is;
(3*) John knows, of a person *s*, that *s* is the person who was the last one to ride the bike;
(4*) John knows, of a reason *r*, that *r* is the reason why the bike was stolen, etc.

It thus seems natural to hold that we should account for ascriptions of practical knowledge in a similar way. Hence,

(1*) John knows, of a way *w*, that *w* is a way of riding a bike.

There are several problems with this strategy. First, one might want to notice that what epistemologists are interested in is knowledge, let it be propositional or practical, not knowledge *ascriptions*. The latter is an interesting topic in the philosophy of language, but it does not necessarily show anything relevant to epistemological concerns regarding allegedly different kinds of knowledge. After all, it is not surprising that if we *call* both propositional and practical knowledge "knowledge" the relevant ascriptions could be reconstructed as having the same logical form—in particular, it is not surprising that the verb "to know" (and its cognates) will govern the same kind of grammatical constructions. That, however, does not show that the *properties*, picked out or ascribed by "knowledge" (and its cognates) in the two locutions and ascribed in knowledge ascriptions, are identical. Nor does it show that the very *concept* expressed in the two cases is the same. It might, but, surely, that

cannot be argued for just by noticing a linguistic analogy. No more than noticing a linguistic analogy could be used to claim that the bank of the river and the Royal Bank of Scotland, say, are the same thing (or different species of the same kind). Indeed, on some contextualist positions, and on several pluralist positions the very same linguistic forms often conceal crucial conceptual and metaphysical differences.

Secondly, one can know a way *w*, which is a way of riding a bike, without thereby being able to ride a bike. Saying, as Stanley and Williamson do, that *w* is a "practical" way of doing *x* does not really solve the problem. For either we are presupposing that those who know *w* have practical knowledge, in which case we are not reducing practical knowledge to propositional knowledge, but we are simply re-describing it, by saying that it consists in *practically knowing* a way *w*, which is a way to ride a bike.[7] Or else, we are saying that those who know this practical way *w*, know it propositionally. As we saw, however, this does not suffice to put one in a position actually to ride a bike.

It follows, then, that practical knowledge is not reducible to propositional knowledge and that the former consists in having certain abilities, which are usually manifested (or that are manifested once the contingent factors that might have inhibited their manifestation are removed).

3 Chomsky: Plato's Problem and Linguistic Competence as Propositional Knowledge

According to Chomsky (1987), language is an exclusive prerogative of our species, is part of our biological endowment and presents only minimal variations among human beings, except for some severe pathologies. According to Chomsky, a person who speaks a language has developed a system of knowledge, associated to certain mental representations, and physically realized in the brain. Moreover, in his view, some aspects of our knowledge and of our understanding are innate. That is, they are part of our biological endowment and are genetically determined, like those aspects of our human nature that determine the fact that we have arms and legs but no wings. In particular, some fixed principles, characteristic of the language faculty, must be attributed to the human organism, as part of its biological endowment. These principles reflect the way in which the human mind functions with respect to the language faculty.

Chomsky, as is well known, opposes any form of behaviorism. In particular, he stands against the idea that to speak and understand a

language is a practical kind of knowledge, similar to riding a bike, and to the idea that the creative aspects of language should be explained in terms of analogies between previously heard sentences and newly produced ones. According to Chomsky, this approach is motivated by anti-mentalistic worries, which depend on an erroneous conception of mentalism and on the mistaken conception that knowledge, in this area, is a kind of competence, ability or skill.

The most important argument put forward by Chomsky concerns the fact that our linguistic abilities may be damaged, for instance after an accident, while our linguistic knowledge remains intact. This is shown, according to him, by the fact that after recovery we speak the same language we spoke before and not a different one. According to him, this shows, further, that we have a series of rules (or principles and parameters) registered at the sub-personal level, which are still there even if we cannot use them and that the acquisition—as opposed to the learning through stimuli and responses—of a language consists in knowing these rules. It consists, for example, in knowing that each well-formed sentence has the subject-verb form and that at least in some languages, such as Italian, the subject's place can be occupied by a "null subject"—that is, a parameter that belongs to the deep syntax of the language but which is not realized in the superficial form of the language (at least not as a separate phonetic entity, since it is usually manifested in the morphology of the verb).

However, we should distinguish between having an ability and manifesting it. A person could still have an ability, but be unable to manifest it for various reasons, as we saw in §2. It is clear, in the case presented by Chomsky, that once the inhibiting factor is removed—e.g. the hematoma that was pressing a part of the brain is absorbed—the subject can manifest her ability to speak her language again, just like a cyclist, who has broken her leg, is able to ride a bike once she recovers from her injury (probably after a bit of rehab and training).

Chomsky anticipates this reply when he observes that, according to common sense, we do not have two concepts of ability, but only one, connected to the fact that the ability is manifested. This is not obvious, however. For it is not clear that, according to common sense, an ability is present just in case it is manifested, or can be manifested upon request. For we may have retained an ability even when it cannot be manifested, particularly when the impediment to the manifestation of the ability is temporary and is removed after a while. Confronted with permanent

impairments, however, our intuitions regarding *both* abilities *and* knowledge of rules seem to change, as we would say that the subject has lost both the ability to speak and her knowledge of the rules of grammar.

Another possible reply to Chomsky may consist in saying that even if common sense sided with him in holding that the concept of ability entails the idea that an ability should be manifested at least upon request, it is not obvious that common sense should rule in this area. Just like other sciences, which are replete with notions that are not in keeping with common sense (e.g. the concept of simultaneity in relativity theory is very different from its commonsensical counterpart), so linguistics too could work with a suitably refined concept of ability. Indeed, Chomsky himself has always insisted that linguistics does not concern itself with language the way common sense, and even philosophy, do. The commonsensical/philosophical notion has it that language is a social, culturally determined object, whose meanings depend on externalist relations—let them be social or causal. By contrast, linguistics, as Chomsky conceives of it, is not concerned with language understood that way (he actually calls it E-language). Rather, it is concerned with I-language. That is to say, an individualistic and internal series of rules[8] that allow us to form and recognize well-formed sentences, together with a series of innate concepts, which allow us to categorize the world as we do.

Either way, we can safely hold that we are not obliged to conclude that the case of the aphasic subject brings grist to the mill of the propositionalist's cause.

4 Language, Rules and Knowledge of Rules

Let us focus on the fact that, according to Chomsky, our linguistic competence is guided by knowledge of the rule that, to be correct, a sentence must have a subject (let it be a null one or otherwise). Clearly, this cannot amount to having explicit knowledge of that rule (or principle), or to being able to make it explicit upon request. As we saw, in this connection we should talk of tacit, or "unconscious" knowledge. The rules (or principles), which supposedly guide us in building and recognizing well-formed sentences, would be written or encoded somewhere, in a place inaccessible to consciousness. As we saw, moreover, an intellectualist will hold that they are operative even if a speaker ignored them, or even denied them.

Yet, in order to even start making sense of this idea one should already buy into the typically cognitivist analogy between minds and computers, where all these rules (or principles) would have to be included in a program we should be born with. However, while with computers we know the program and know how it gets into the machine, when it comes to us we know very little both about the program's structure and about its actual provenance. Take the "principles and parameters" version of universal grammar. Only a subpart of it would be innate. For the exposure to one's own language would be necessary in order to activate the relevant parameters whenever appropriate. For instance, a speaker of Italian would have to be exposed to that language in order to activate the parameter of the null subject, while a speaker of English or French, by being exposed to her language, would never activate it. Notice, however, that once you start admitting that the formation and the recognition of grammatically correct sentences are possible only once exposure to one's language has taken place, it becomes dubious that no actual learning is involved.[9] Chomsky would have us believe that the exposure to one's language would just trigger a parameter, like exposure to a given kind of food, liquid or air composition might trigger this or that chemical reaction in our metabolism. Yet, this analogy seems to be motivated by an anti-"behaviorist" prejudice in its turn. It is as if Chomsky could not accept the idea that what happens is that we expose and train children to the use of language for a considerable amount of time, until they are actually able to form well-formed sentences, by the lights' of their own respective languages. Maybe this anti-behaviorist prejudice would be motivated if we thought of training and learning as a kind of Pavlovian conditioning, but that is not what happens in reality. We do not habituate children with punishments and rewards to always react a certain way, or simply to repeat sentences uttered by the adults. We will presently consider a more credible description of what happens at those stages of early life.

For now, it is worth remarking that whereas the analogy between minds and computers might seem convincing insofar as minds would be the equivalent of programs, as soon as we go one level down, we are only left with brains and neuron firings. The idea that, at that level, there should be rules guiding our behavior is problematical. Rather, there are neural circuits that somehow follow certain patterns once so habituated, thanks not just to exposure to one's language but to actual training. Moreover, as is well known, once brain damage has occurred,

certain cognitive tasks might be taken over by different sets of neurons. True, one might say that the physical realization of that piece of program would have changed, while the program would still be the same. Yet, clearly, this would mean conflating the fact that we can describe the task in similar or identical ways, at a suitably abstract level, with saying that the same piece of program—that is, that very rule—is actually operative somewhere in our brains.

Yet, if we renounce the idea that the relevant rules are encoded in our brains and guide our linguistic performance, in favor of the idea that they are individuated *ex post*, to describe an ability underwritten by a given neurological structure and functional activity, the very idea that linguistic competence is a case of propositional knowledge of rules evaporates. For, then, that knowledge is neither explicit nor tacit. Moreover, the very rules, which should be its contents and that each of us should have stored in her own mind, are simply the rules that linguists come up with to model our linguistic abilities, at an extremely high level of generality and abstractness, in the light of the data acquired by investigating the enormous variety of human languages.[10]

5 Innateness and Empirical Data

Let us now turn to Chomsky's empirical arguments in favor of innatism and of the idea that linguistic competence consists in propositional knowledge of a series of syntactical rules (or principles), of a set of innate concepts and of a series of phonological rules. He thinks that crucial empirical data could not be explained unless we accepted those hypotheses. However, like any other inference to the best explanation, even Chomsky's does not lead to sure-fired conclusions and, actually, some of the data are not as solid as Chomsky holds.

The first and most important empirical evidence is the fact that children quickly acquire language and are able to form and recognize well-formed sentences, despite the "poverty of the stimulus". That is to say, children are exposed only to a limited number of sentences. Nonetheless, they quickly become able to form new sentences they never heard before.[11]

According to Chomsky, another indicative factor that the language faculty, which is the program that universal grammar describes, is innate and common to the human species as a whole is that even subjects affected by serious syndromes, like Down syndrome, acquire language and so do children who are blind from birth.

In addition, according to Chomsky, the speed of acquisition of vocabulary in children does not allow for alternatives to the idea that they possess a range of innate concepts and that all they need is to learn how to label them in their respective languages. Conceptual innateness, moreover, explains the fact that definitions can be useful, despite their being imprecise, according to him. Furthermore, Chomsky maintains that these basic concepts (such as physical object, human intention, will, cause, ends, etc.) form a hierarchy and are compositional.

It should be noted, however, that all these data are not incompatible with anti-innatism and with the idea that to speak a language is a practical kind of knowledge.

Regarding the speed of acquisition of language and the poverty of stimulus, it is evident to anyone who has ever actually interacted with children that the time of acquisition of language, from a phonological, syntactical and semantic point of view, is quite long. Only around age 3 do children start forming simple well-formed sentences. The same goes for mastery of the phonological aspects of their language, as well as for mastery of a reasonably wide vocabulary. Chomsky often compares language acquisition with number acquisition but, in the latter case, things are worse still. At age 3, counting is out of the question and while children seem to grasp the meaning of 'one' and possibly 'two', they seem to lump everything together as 'many' or 'more' from 'three' onwards. What they can do, like several other animals, is rather to discern ratios between aggregates. Moreover, if they utter further number words, perhaps because they have heard them from parents and older siblings, they clearly have no grasp of what they mean. That is decisive evidence of the fact that these words are not labels for already possessed concepts. Rather, they are merely linguistic placeholders for concepts, which are acquired much later. Nor is it credible to suppose that we should have only a very limited amount of innate number concepts, while the rest of them would be acquired later.

Concerning the idea that certain fundamental concepts are innate and that what is learnt are merely labels to name them, a proper treatment of the issue would deserve a separate chapter. I will simply mention Tyler Burge's important contributions regarding perception, subitizing and aggregates' ratio discrimination.[12] They all go in the direction of showing that, in order to describe those capacities, we need not ascribe any concept such as the previously listed ones, or indeed numerical concepts. Rather than attribute them concepts, which should rather be identified

and attributed on the basis of quite complex inferential abilities, we can explain their purposive behavior by positing perceptual attributives—that is, roughly, purely perceptual representations, with a compositional structure. Hence, there is actually no need to embrace conceptual innatism.

Again, the fact that language is acquired also by subjects affected by Down syndrome or blindness since birth clearly shows that human beings have a biologically determined neurological structure that enables them to acquire language. It shows, furthermore, that that is the case also for subjects affected by serious syndromes or impairments. Yet, this falls short of proving that we also have an innate set of rules (or principles) somehow encoded in our brains.

Moreover, not all pathologies are compatible with language acquisition. Deaf-muteness requires a special training concerning the use of sign language. Hence, this pathology, while not incompatible with language acquisition as such, is incompatible with the standard way in which human beings acquire language. This shows, once more, how the acquisition of language is possible only if, on top of having certain neurological functions, one is actually exposed, and trained to the use of the relevant symbolic system. A given neurological structure is thus necessary but not sufficient to acquire a language. In addition, it may be that that neurological structure can perform a certain function only if the relevant training takes place within a certain age. Yet, this is entirely compatible with the fact that speaking a language is, at bottom, a practical kind of knowledge.

Moreover, nothing prevents us from thinking that that exposure and that training allow us to *learn* a language—rather than merely develop some aspects of a "language faculty". Furthermore, nothing precludes supposing that rather than be born with a given set of linguistic rules (or principles), that exposure and that training are actually necessary in order to enable our brains to work in the way required to master a language.

Let me close with some positive suggestions regarding language learning and the idea that speaking a language consists in a rule-governed practice. We can find inspiration on these fronts once again in the Wittgenstein-Ryle view. Of course, central to that conception is the idea of language as a social phenomenon, whose rules are determined by shared and repeated patterns of use among members of the linguistic community. We have already seen why Chomsky's innatism is far from obvious and so, in these concluding remarks, I will put it on a side, at least for the sake of argument. When we say that speaking a language is a

rule-governed practice, however, we have to be clear about the nature of these rules and how they can be said to govern our practice. For we do not want paradoxically to end up defending the propositional account of knowledge of rules Chomsky put forward, which is detachable from his endorsement of innatism.

As we anticipated (in §1), linguistic rules are established by use, in the Wittgenstein-Ryle picture, and key to that conception of language learning is the idea that rules are acquired purely practically. Once one possesses a language, one can reflect on the practice and formulate the rule, or understand its formulation by other members of one's linguistic community. Indeed, there can be intermediate phases, in which subjects have some grasp of certain linguistic rules and they themselves correct supposedly wrong applications of those rules, or test them by applying the rules to new cases to which the latter may or may not apply. However, this requires mastery of substantial chunks of language already. Hence, the idea of an explicit (or semi-explicit) learning of rules cannot be used to explain the initial phases of language acquisition. At those stages, children try, stumble, fall and sometimes succeed and, little by little, they acquiesce in a practice. Even the correction by an adult need not be understood as passing a rule on to them. That is, at those stages, an instinctive behavior, like saying 'goed' or 'taked' is replaced by a culturally determined one that prompts them to say 'gone' or 'taken'. What is considered to be the correct practice—that is, the one in keeping with the rest of one's linguistic community—is inculcated, rather than taught as an explicit rule.[13] Nor is it necessary that at later stages there actually be anything like an explicit formulation of linguistic rules, made by the subject herself or by other members of her community, in order for her to be able to participate in the relevant linguistic practice. By being exposed to that practice subjects learn how to take part in it, little by little and through a complex training which mainly consists in talking to children in a variety of contextual situations, which are neither fixed nor predetermined, even though they are very often recurrent. Children pick up some basic words and use them to participate in the relevant activities, calling for their primary care-givers, or for expressing their basic needs, or for taking part in socially determined practices such as greeting, playing simple games, etc. They then become able to form sentences and, after a while, they start doing so in a creative way, which manifests the fact that they have practically grasped the fact that language is often (not always!) compositional. Yet again, this does not mean that

they have propositional knowledge of a rule, however tacit that knowledge might be, let alone that the rule is encoded in their minds/brains. Rather, they have learnt to behave in ways that tally with the rest of the relevant community's usage and that, by being accepted, are certified as correct by some of its members. Hence, linguistic competence is primarily a practical kind of knowledge—it is a know-how, rather than a know-that—which consists in being able to behave in conformity with the rest of our linguistic community's practice.[14] Such a practice is shared and constant in time, though it may be subject to changes. Due to its regularity, and to the fact that deviant behavior tends to be corrected, that practice can be seen as governed by rules. Once one has acquired a language, one can actually reflect on the linguistic practice itself and arrive at an explicit formulation of at least some of its rules. Yet, as we saw, theoretical knowledge of how we do certain things is neither necessary nor sufficient for being able to do them. It is this ability we are trying to characterize when we inquire into the nature of practical knowledge in general, and of linguistic knowledge in particular.

The interesting and promising aspect of current psychological research in this area is—perhaps not surprisingly—that it is going past the Chomskyan model and is vindicating the core aspects of the picture of language, language acquisition and linguistic knowledge actually put forward by Wittgenstein and Ryle long ago.

For instance, statistical learning accounts show that children systematize the language they are exposed to based on the frequency of forms, rather than on the basis of any previous knowledge of syntax or of innate concepts. As Moyal-Sharrock (2016, pp. 8–10) aptly summarizes these findings: "they generalize from cues, not from rules". This explains easily why they tend to add the suffix '-ed' to irregular verbs, but also why, as Melodie Dye (2010) puts it, "they end up homing in on and reproducing only the most frequent patterns in what they hear. In doing so they fail to learn many of the (…) idiosyncrasies present in adult speech".

Moreover, the "usage-based linguistics", proposed by Michael Tomasello and others, is collecting copious empirical data that go in the direction of denying the innateness of language, and the existence of a dedicated language faculty, with very abstract algebraic rules, that would constitute (one version or the other) of Chomsky's universal grammar. Rather, these studies support the hypothesis that grammar is the product of history, which has evolved in numerous different ways. They also support the idea that language is learnt by being immersed

in a practice and that such learning draws on several cognitive faculties, which have specifically nothing to do with language, such as categorization, "mind-reading" and analogy making. In particular, according to Tomasello, grammar is something children discern in the various actual sentences they are exposed to. In keeping with the "statistical learning" approach, Tomasello holds that, given the potentially infinite number of meaningful yet ungrammatical generalizations children could make, they appear to home in the correct ones because "they are sensitive to the fact that the language community to which they belong conforms to one [specific] norm and communicates an idea in just 'this way'" (Tomasello and Ibbotson 2016, p. 17). Coordination is, after all, conducive to successful communication.

In a more Wittgensteinian and Rylean spirit, we may say that children learn linguistically to behave as the rest of their community does. This allows us to describe them as followers of the very rules that, in its turn, their linguistic community can be said to share, in virtue of behaving the way it does, and in virtue of keeping the practice stable by correcting subjects who deviate from it, to enforce the rule.

Of course, when philosophical pictures are passed on to science they become subject to empirical (dis)confirmation. Yet, after years of oblivion, the Wittgenstein-Ryle picture, broadly conceived, is now considered a serious contender in the scientific arena. Let time decide which ones of these two rival paradigms will prevail.

6 Conclusions

In this chapter, we have seen how, contrary to what Stanley and Williamson hold, practical knowledge is neither reducible to propositional knowledge, nor to be considered a species of it. We have also seen how Chomsky's theory, according to which linguistic competence consists in propositional knowledge, whose content is a set of rules (or principles), is quite objectionable, both for theoretical reasons, and for considerations of empirical adequacy. In particular, Chomsky's arguments are dubious and the data that he thinks can only be explained by accepting innatism and the idea that we would propositionally know a system of rules (or principles) are not always solid. Finally, when they are, they can perfectly well be explained by renouncing both innatism and the idea that we have propositional knowledge of the rules (or principles) of universal grammar.

Notes

1. Dummett's views developed over the years, as he abandoned the initial claim that knowledge of language is implicit. However, his considered view, shared by Eva, is at odds with the interpretation of Wittgenstein's position I propose in the following.
2. See Ryle (1949, pp. 46, 55).
3. I think missing this crucial distinction is what leads several scholars to denying that know how is an ability. See for instance Stanley and Williamson's quick dismissal of that idea (2001, p. 416). Of course, theoretical knowledge of how to do something may be an interesting research topic in its own right. Yet, once the difference with its practical counterpart is clearly in view, there is no temptation to think of the latter as propositional, while it becomes trivially obvious that the former is just an instance of propositional knowledge.
4. See Wallis (2008, p. 140, quoted in Fantl 2012).
5. I prefer to avoid calling "implicit" the kind of knowledge that could be made explicit, even if normally it isn't. The real contrast is between explicit or explicit-able knowledge on the one hand, and tacit or unconscious knowledge on the other.
6. See Wallis (2008, p. 140).
7. Stanley and Williamson actually deny that their aim is to reduce practical knowledge to propositional knowledge, but then one wonders what the interest of their proposal would be if it were just a re-description of practical knowledge, or a proposal regarding simply the logical form of practical knowledge ascriptions (cf. Stanley and Williamson 2001, pp. 433–434). Indeed, they say that their aim is to show that knowledge how is a "species" of knowledge that. Let us grant them that there is a significant difference between this claim and the thesis that knowledge how reduces to knowledge that. Still, for the reasons given in the main text, I do not think they have succeeded in showing that much either.
8. Or of principles and parameters, or of whatever the most recent development of generative grammar posits as characteristic of this internal program.
9. For a very useful survey of the empirical data in this connection, see Moyal-Sharrock (2016, p. 15, fn. 24 and pp. 21–23).
10. Indeed, the very idea of an innate structure that imposes universals has been rejected from a biological perspective. For a useful survey of the empirical work in this area, see Moyal-Sharrock (2016, pp. 14–16).
11. As Moyal-Sharrock (2016, p. 4, fn. 3) points out, Chomsky is no longer concerned by the degeneracy of the data. For empirical studies have shown that speech addressed to children is highly regular. Roughly, only 1 out of 1500 utterances addressed to children is ungrammatical.

12. See Burge (2010). Cf. also Bermúdez (1998). I have discussed Burge's position at length in Coliva (2012). See also Panza and Coliva (2018) for a discussion of numerical cognition.
13. Indeed, Wittgenstein talks of "drilling" and "training", as opposed to "explaining" in this connection. Cf. Wittgenstein (1953, §5).
14. Of course, the linguistic competence and language learning for which this picture makes sense are the ones relative to one's mother tongue(s). The acquisition of a second language involves many explicit formulations of syntactical and semantic rules (as well as of phonological ones). It also involves a lot of inference to the best explanation while being exposed to native speakers of that language. However, the acquisition of a second language is more like a translation of one language into another one than learning a language *tout court*.

References

Bermúdez, J. 1998. *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Burge, T. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

Chomsky, N. 1987. *Language and Problems of Knowledge*. Cambridge, MA: The MIT Press.

Coliva, A. 2012. Critical Notice of Tyler Burge's *Origins of Objectivity*. *Disputatio* 4 (33): 515–530.

Dummett, M. 1981. Objections to Chomsky. *London Review of Books* 3 (16): 5–6.

Dye, M. 2010. The Advantages of Being Helpless: Human Brains Are Slow to Develop—A Secret, Perhaps, of Our Success. *Scientific American*, Feb. https://www.scientificamerican.com/article/advantages-of-helpless/.

Fantl, J. 2012. Knowledge How. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/knowledge-how/.

Fodor, J. 1968. The Appeal to Tacit Knowledge in Psychological Explanation. *The Journal of Philosophy* 65 (20): 627–640.

Moyal-Sharrock, D. 2016. Universal Grammar: Wittgenstein Versus Chomsky. In *A Companion to Wittgenstein on Education: Pedagogical Investigations*, ed. M.A. Peters and J. Stickney, 1–33. Dordrecht-Kluwer: Springer.

Panza, M., and A. Coliva. 2018. Is Numerical Cognition Cognition of Natural Numbers? *Synthese*, forthcoming.

Picardi, E. 1997. Is Language a Natural Object? In *Thought and Ontology*, ed. M. Sainsbury, 107–123. Franco Angeli: Milano.

Picardi, E. 2001. Normativity and Meaning. In *The Dialogue. Proceedings of the 2nd Meeting Italian/American Philosophy 1999*, ed. R. Dottori, 233–250. Hamburg-London: Verlag.

Ryle, G. 1949. *The Concept of Mind*. Chicago: The University of Chicago Press.
Ryle, G. 1971. Knowing How and Knowing That. In *collected Papers*, vol. 2, 212–225. New York: Barnes and Nobles.
Stanley, J., and T. Williamson. 2001. Knowing How. *The Journal of Philosophy* 98 (8): 411–444.
Tomasello, M., and P. Ibbotson. 2016. Evidence Rebuts Chomsky's Theory of Language Learning. *Scientific American*, Sept 7. http://www.scientificamerican.com/article/evidence-rebuts-chomsky-s-theory-of-language-learning/.
Wallis, C. 2008. Consciousness, Context, and Know-How. *Synthese* 160: 123–153.
Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.
Wittgenstein, L. 1969. *On Certainty*. Oxford: Blackwell.

A Plague on All Your Houses: Some Reflections on the Variable Behaviour of "Knows"

Crispin Wright

It is a great pleasure to contribute to this volume in celebration of the life and work of Eva Picardi, and I am most grateful to the editors, Annalisa Coliva, Paolo Leonardi, and Sebastiano Moruzzi, for providing me with the opportunity to do so. My own personal acquaintance with Eva dates back to the 1990s and a series of delightful Summer Schools for graduate students held at various Northern Italian universities in which we both participated. We had each been former students of Sir Michael Dummett and had absorbed from him a deep interest in Frege and in philosophy of language in the Fregean tradition that Dummett did so much to foster. I was greatly impressed by the depth of Eva's Frege scholarship—as well as by her very forceful philosophical personality!—and was delighted when she later agreed to work as a consultant with the team, then based at the *Arché* centre in St Andrews, that was inching towards completion of the first unabridged English translation of Frege's *Grundgesetze* and to which, though already struggling

C. Wright (✉)
New York University, New York, NY, USA
e-mail: cw43@nyu.edu

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_17

with failing health, Eva made very valued contributions as the work neared completion.

My chapter here, though, is not about our shared Fregean concerns. Eva took a keen interest in the 'linguistic turn' taken by the contemporary discussions of relativism and was, I think, both impressed by and suspicious of it. It is my hope that she would have found something of interest in what follows.[1]

1 The Variabilist Reaction against Traditional Epistemology

It is fair to say that from the time of the *Theaetetus* until relatively recently, theorists of knowledge tended to conceive their central task as being to explain in what knowledge consists; more exactly, to explain what further conditions need to be satisfied by a true belief if it is to count as knowledgeable. The widely accepted failure of the post-Gettier debates to execute this task convincingly has motivated a very different tendency in mainstream contemporary epistemology. This tendency, influentially promoted by Timothy Williamson in particular, is *epistemic primitivism*: to concede that knowledge is, as Williamson puts it, 'prime'—that it is a fundamental, irreducible cognitive relation. Knowledge, on the primitivist view, is a basic epistemological kind, and to know is to be in a basic, *sui generis* attitudinal state. There can therefore be no correct analysis of it in terms of other, supposedly constitutive or more fundamental cognitive states (true belief + X). The post-Gettier "X knows that P if and only if…" cottage industry was doomed to disappointment for this reason. To the contrary, for the epistemic primitivist, it is in terms of knowledge that other epistemic notions—justification, evidence, warranted assertion and rational action—are to be understood.[2]

This primitivism, however, shares three traditional assumptions with the reductionism it is set against. They can be wrapped together as the compound idea that knowledge is a *determinate, objective, purely cognitive* type of condition—hence something at which the aspiration of reductive analysis could be sensibly (even if mis-) directed. If we unpack that, however, we find the following three distinct thoughts:

– First, ascriptions of knowledge, that X knows that P, are *contentually invariant* as far as the semantic contribution of 'knows' is

concerned. More specifically, once the referent of 'X', the identity of the proposition that P and the time reference associated with 'knows' are settled, the result is a unique proposition, the same for any competent thinker who considers it.
- Second (although this would normally be taken to be entailed by the first point) this unique proposition has one and the same truth-value, no matter who asserts or assesses it.
- Third, this truth-value is determined purely by the cognitive achievements of the subject, irrespective of what else, other than that part of her total information relevant to the judgement that P, is true of X. In particular, such aspects as X's (or anyone else's) *interest* in whether P is true, or *what is at stake* for her in its truth, or the range and specifics of counter-possibilities to P that occur, or are *salient*, to X—in short: such, as they are often described, "non-traditional" or as I shall say *pragmatic* factors—have no bearing on the matter.

The recent tendency that provides the subject matter of this chapter is the rejection of one or more of these traditional assumptions in favour of one or another form of *variabilism*: broadly, the notion that whether an ascription of knowledge may correctly be regarded as true may depend on pragmatic factors that pertain to the circumstances of the ascriber, or to those of a third party assessing the ascription, or on pragmatic aspects of the circumstances of the ascribee. Although well short of a consensus, there has developed a considerable body of opinion that agrees that *some* form of epistemic variabilism is called for if justice is to be done to the actual employment of "knows" and its cognates.

In what follows, I will review one kind of consideration that has been taken to support that view, critically compare and assess some of the resulting variabilist proposals, and recommend a conclusion both about them and about the prospects for primitivism.

2 The "Data"

Probably the most influential motive for variabilism about "knows" draws on a range of putative linguistic 'intuitions' concerning proprieties of knowledge-ascription provoked, at least among many of the philosophers who think about such things, by imaginary cases of a kind first put forward by Stewart Cohen and Keith DeRose.[3] We can illustrate

by reference to a version of DeRose's famous Bank Case. Suppose it is Friday afternoon, and Ashley and Bobbie are considering whether to bank their salary cheques. There are long queues at all the bank counters. Ashley recalls being at the bank on a Saturday morning two weeks ago and says, "Let's come back tomorrow. **I know the bank will be open tomorrow morning**." Suppose that the bank will indeed be open on the Saturday morning.

Case 1 (*Low stakes*): Suppose that there is no particular reason to ensure that the cheques are banked sooner rather than later—say, by the following Monday. Then
Invited intuition: Ashley's recollection of Saturday morning opening two weeks ago suffices for her to speak truly.

Contrast that scenario with
Case 2 (*High stakes*): The couple's mortgage lender will foreclose unless the cheques are in the account by Monday to service their monthly repayment. Ashley and Bobbie know this. Bobbie says, "But what if the bank has changed its opening hours? Or what if the Saturday morning opening was some kind of one-off promotion?" Ashley says, "You're right. **I suppose I don't really *know* that the bank will be open tomorrow** (even though I am pretty confident that it will). We had better join the queue."
Invited intuition: Again, Ashley speaks truly. There is too much at stake to take the risk of e.g. a change in banking hours.

So the suggested conclusion is that "I know the bank will be open tomorrow" uttered by Ashley is true in Case 1 and false in Case 2 even though all that is different between the two are the costs to Ashley and Bobbie of Ashley's being wrong. Only pragmatic factors have changed. Everything that might be mentioned in a traditional account of knowledge—as we would naturally say, all Ashley's relevant evidence or information—remains the same.

Two further cases may seem to prompt another important conclusion:

Case 3 (*Unknowing high stakes*): The couple's mortgage lender will indeed foreclose unless the cheques are in the account by Monday to service their monthly re-payment but Ashley and Bobbie are unaware of this (they habitually leave what looks like circular mail from the mortgage

company unopened and have missed the reminder). The dialogue proceeds as first described above, with Ashley asserting, "**I know the bank will be open tomorrow morning**".
Invited intuition: This time, Ashley speaks falsely.

Compare that with
Case 4 (*Unknowing low stakes*): Ashley and Bobbie actually have no good reason to ensure that the cheques are banked before Monday but, misremembering the notice from the mortgage company, they *falsely* believe that Monday will be too late. The dialogue proceeds as in Case 2.
Invited intuition: This time Ashley's disclaimer, "**I suppose I don't really know that the bank will be open tomorrow**" is false.

The suggested conclusion from cases 3 and 4 is this: that when changes in pragmatic factors convert a true knowledge-ascription into a false one, or vice versa, it is *actual* changes that matter, rather than thinkers' impressions of what changes in such factors may have taken place.

3 Varieties of Variabilism

How to explain these 'data'? The space of theoretical options will include at least three quite different kinds of proposal: one for each of the traditional assumptions distinguished in section I. First, we might propose that although knowledge-ascriptions are contentually invariant (in the sense there specified), the proposition thereby expressed may take different truth-values in different circumstances, depending on variation in the pragmatic factors applying to its *subject*, X. This is the thesis, proposed separately by Stanley and Hawthorne,[4] that is most often termed *interest-relative invariantism* (IRI).[5] The details of a proposal of this kind will naturally depend on just what kinds of pragmatic factor are deemed relevant—variation in *what is at stake* is what seems germane in the various scenarios in the Bank Case. IRI allows, apparently, that a pair of subjects may both truly believe that P on the basis of the same evidence or cognitive achievements yet one know that P and the other fail to know that P if they suitably differ in pragmatic respects. I'll come back to this.

Second, we might hypothesise that the variability in truth-value of knowledge-ascriptions across the kinds of situation illustrated is actually a product of variation in *content*. The specific version of this proposal made by DeRose and Cohen is standardly termed *ascriber contextualism*

(henceforward simply "contextualism"). In its original and basic form, this view holds that the (level of) cognitive achievement that is required of X by the truth of an utterance of "X knows that P" varies as a function of pragmatic aspects—needs, stakes, saliences—*of the speaker*. Thus in an example like the Bank Case, variation in pragmatic aspects of a *self*-ascriber across actual, or hypothetical, cases may result in (actual, or hypothetical) tokenings of "I know that P" demanding different—more or less exigent—levels of cognitive achievement if they are to count as true. The truth-conditions, hence content, of tokens of such an ascription can vary, even though the only differences in their respective contexts of utterance pertain to the situation of the speaker in purely pragmatic respects.

The third option—that of *knowledge relativism*, fashioned on the model of assessment-sensitivity as developed by John MacFarlane[6]—shifts the location of the pragmatic factors once again, this time to anyone who evaluates a knowledge-ascription, whether or not they are its original author. So a single token of 'X knows that P' may properly be assigned different truth-values in differing contexts of assessment, whether or not distinct assessors are involved, depending on the situation in pragmatic respects of the assessor. Thus Ashley may again quite correctly return different verdicts on a self-ascription of knowledge that the bank will open on the Saturday in the two contexts described.

An alert reader will have noted that these three types of variabilist view exhibit disagreement in two dimensions. Agreeing that the truth-value of a knowledge-ascription may vary as an effect of variation in non-traditional pragmatic factors, they disagree about the *location*—subject, ascriber, or assessor—of the relevant factors; but they also disagree about the *semantic significance* of such variation. For both knowledge relativism and interest-relative invariantism, variation in pragmatic factors is of no semantic significance at all: rather, one and the same proposition gets to vary in truth-value in tandem with variation in the pragmatic characteristics of the subject, or assessors of that proposition. For knowledge contextualism, by contrast, at least in its classic form, it is the proposition expressed by a particular knowledge-ascription that varies in a fashion sensitive to the pragmatic factors. Ashley's tokens of "I know the bank will be open tomorrow morning" express different propositions in the low-stakes and high-stakes scenarios outlined. There is therefore conceptual space for three further types of view that are the duals in these two dimensions of the three distinguished. There is, first, scope for a

kind of contextualism—an instance of *non-indexical* contextualism[7]—that agrees with classical contextualism on the matter of location but disagrees on the matter of semantic significance. On this view, Ashley's two imaginary tokens of "I know the bank will be open tomorrow morning" express the same proposition in the low-stakes and high-stakes scenarios, but this proposition takes a different truth-value as a function of the difference in what is at stake for the ascriber—Ashley—in those scenarios. Second, there is scope for a view which, like classical contextualism, regards ascriptions of knowledge as varying in their content (truth-conditions) as a function of variation in pragmatic characteristics but holds, like interest-relative invariantism, that the relevant characteristics are those not of the ascriber but of the subject, or subjects, to whom knowledge is ascribed. On such a view, a predicate of the form, " … knows that P", will vary in its satisfaction-conditions rather as e.g. "… is sharp enough" so varies depending on whether it is being applied to a bread knife or a surgical scalpel. And finally, there is scope for an example of the view that *content itself* is, locally, assessment-sensitive: that what proposition is expressed by a token knowledge-ascription is itself a function of pragmatic characteristics of an assessor of it, with assessment-sensitivity of truth-value merely a consequence of such assessment-relativity of what is said.[8] I do not know if anyone has ever seriously proposed a view of either of these two latter kinds for the semantics of "knows" but in any case neither will feature further in the discussion to follow. However in view of the difficulties, to be touched on below, that classical contextualists have encountered in trying to make good the claim that "knows" is indeed semantically context-sensitive, its non-indexical counterpart presents as worthy of serious consideration. It will surface from time to time below.

4 The Location Question

So, *whose* standards (saliences, interests, etc.) count? The cases 1-4 considered to this point involve *self*-ascriptions of knowledge. So they have the subject of the knowledge-ascription coincide with the ascriber coincide with an assessor. They therefore can suggest at most that we should be receptive to *some* sort of variabilism. They are powerless to motivate one rather than another of the variabilist views. Can we find some crucial experiments?

Here is a simple kind of case that has seemed to contextualists to favour their view over IRI:

Case 5 (*High-stakes ascriber, low-stakes subject*): Ashley and Bobbie are situated as in Case 2. They ask Chris, another customer who is leaving the building, whether the bank will be open tomorrow. Chris says "Yes, I happen to know it will—I was in here a couple of weeks ago on a Saturday." Ashley says to Bobbie *sotto voce*, "Hmm. **That chap doesn't know any better than we do**. We had better join the queue."
Invited intuition: Ashley speaks truly even though—as we may suppose—there is nothing at stake for Chris, the subject, in whether the bank will open on the Saturday or not. Here, it seems the interests that count are those of the ascriber, even when the subject is someone else whose interests are different (and less urgent).

The significance of this kind of case is prima facie countered, however, by the following simple case that may seem to point back towards IRI:

Case 6 (*Low-stakes ascriber, high-stakes subject*): Ashley, Bobbie and Chris are again situated as in Case 5. Chris is puzzled that Ashley and Bobbie have joined the queue again notwithstanding the advice they were just given about Saturday opening and asks them about this. They explain their concern about the risk of foreclosure of their mortgage. Chris says, "OK, I understand now. I guess you guys had better not assume that the bank *will* be open tomorrow."
Invited intuition: Chris speaks truly. But since "You know that P but had better not assume that P" is some kind of conceptual solecism, Chris's remark is presumably a commitment to "**You do not know that the bank will be open tomorrow**."[9]

So, neither contextualism nor IRI does well in all cases—in fact they do just as well and badly as each other: well enough in cases where subject and ascriber are identified, but badly in various kinds of case where they are distinct—which are of course the crucial cases. This might encourage the thought that *both* have the location issue wrong and one might therefore wonder whether knowledge relativism promises an overall better ride. And indeed we can very simply modify Case 5 to get one that seems to favour knowledge relativism over contextualism *and* IRI:

*Case 5** Ashley and Bobbie are dithering in the foyer and then merely overhear Chris (in a phone conversation) say "Look, I don't need to wait here now. My partner, Denny, was here a couple of weeks ago on a

Saturday and can vouch that this bank will be open tomorrow." Ashley remarks, *sotto voce*, "We can't rely on that; **that Denny doesn't know any better than we do**."
Invited intuition: Ashley speaks truly.

However while knowledge relativism may possibly best explain some intuitions in cases like this where subject, ascriber and assessor are all distinct, it faces the basic problem that it must coincide in its predictions with contextualism in any case where ascriber and assessor are one. So any two-agent problem cases for contextualism, like Case 6, are problems for relativism too.

These conflicting intuitions present a potential paradox if we think that they do, near enough, show that there is *some* kind of relativity to pragmatic factors in the offing. How can that be so if the intuitions also suggest that each of the possible hypotheses about location is open to counterexample?

5 A Contextualist Attempt to Explain Away the Recalcitrant "Data"

There is a response that at least one leading contextualist has offered that is potentially something of a game-changer. Keith DeRose observes[10] that in taking patterns of conversation like those illustrated by Case 6 to constitute prima facie counterexamples to contextualism, we are implicitly taking it for granted that the mechanism whereby the context of a token knowledge-ascription contrives to set the standards for its truth is simply by identifying them with the standards of the ascriber: that "X knows that P" as uttered by Y is true just if X's relevant epistemic situation, replicated by Y but without change in the pragmatic aspects of Y's situation, would suffice for the truth of "Y knows that P". DeRose points out that there is absolutely no reason why that has to be the only kind of case. It is very familiar that in a wide range of examples—'impure indexicals' like some personal pronouns, demonstratives, and gradable adjectives—the semantic values of context-sensitive expressions featuring in particular utterances are settled as a function, in part, of the intentions of the utterer. It is therefore open to the contextualist to allow a similar role for the intentions of the author of a knowledge-ascription in determining the standard of epistemic achievement to be applied in fixing its

truth-conditions. This can of course be the standard she would (take herself to) have to meet in order to satisfy the relevant ascription. But it need not be. In certain contexts—like that of Case 6—an ascriber may instead set a standard that defers to the needs, interest, or saliences of the subject. In such a case, IRI and contextualism will coincide in their predictions of the truth-conditions of the knowledge-ascription.

I described this 'flexible contextualist' manoeuvre as a potential game-changer. It is, of course, merely ad hoc unless a principled and comprehensive account is provided of the conditions under which relevant variations in a speaker's intentions can be expected, enabling empirically testable predictions of variable truth-conditions. DeRose expends some effort in that direction, to not implausible effect. His basic suggestion is that knowledge-ascriptions may be harnessed to two quite different kinds of project: whether X knows that P may be of interest because one wishes to rate X as a potential *source of information*; but it may also be of interest in the context of assessing X's performance as a *rational agent*. In the former type of case one will naturally impose standards on X's claim to knowledge appropriate to one's own needs and interests. Just this is what seems to be happening in the kind of high-stakes ascriber, low-stakes subject cases illustrated. But in the latter type of case, when the focus shifts to what it is rational for X to do, it may well be (one's conception of) X's needs and interests that determine what level of cognitive achievement it is reasonable to demand if X is to be credited with the knowledge that P. And this seems to be the driver for the (invited) intuitions operative in kind of the low-stakes ascriber, high-stakes subject cases like 6.

I have no space here to consider further whether the flexible contextualist manoeuvre can be developed so as to deliver fully satisfyingly on its initial promise. However two points are worth emphasis. The first is that an exactly analogous flexibility on the location question is, obviously, available to knowledge relativism. Whatever potential shifts of interest are offered to explain variations in the location of standards from the point of view of a knowledge-ascriber, they will be available also to explain such variations from the perspective of a knowledge-ascription assessor. Flexibility thus offers no prospect of an advantage for contextualism over relativism. Second, there is no analogous move open to IRI, which is stuck with the idea that the standards for the truth of a knowledge-ascription are inflexibly set as a function of the needs, interests, or saliences of its subject. If IRI is to restore dialectical parity after (and presuming the success of) the flexible contextualist manoeuvre, it

must therefore explain away cases, like Case 5, where the location seems to go with an ascriber (or assessor), rather than the subject, as some kind of linguistic mistake. What are the prospects?

It is important to take the full measure of the challenge. Any presumed *knowledgeable* ascription of knowledge to a third party entails—by closure and factivity—an ascription of the same knowledge to oneself. And of course if IRI is right, and one's standards are relatively high, one may not have that knowledge. In that case, one won't be in position to ascribe it to a third party either, whatever their standards. There is therefore in general no difficulty for IRI in explaining our *reluctance to ascribe* knowledge in such cases. That, however, is not the relevant *explanandum*. What the defender of IRI has to explain—what the high-stakes ascriber low-stakes subject examples are meant to illustrate—is a readiness of high-stakes ascribers to (falsely) *deny* knowledge that P to a relevant low-stakes subject. (Thus Ashley: "That person doesn't know any better than we do.")

It would take us too far afield to pursue the details of all the responses that defenders of IRI have offered to this challenge. Let me here merely record the opinion that they have not so far proved successful.[11]

6 Ugly Conjunctions

We have so far been concerned with the challenge to the different variabilist views to capture and explain not just some but all the pragmatically variable patterns of use of "knows" and its cognates that, according to the 'intuitions', competent speakers seem to find acceptable. And at this point, provided they are prepared to go 'flexible', and thus steal the cases that otherwise favour IRI, contextualism and relativism seem to be tied in the lead. But there is also an obverse challenge: to avoid predicting uses to be acceptable which are apt to impress as anything but. How do the different theories fare on this?

IRI imposes a condition on knowledge-ascriptions as follows:

> X knows that P at t is true only if X's belief at t that P is based on cognitive accomplishments that meet standards appropriate to X's practical interests (or whatever) at t,

and consequently appears to do very badly. Suppose X fails this condition—his practical interests are such that it is vitally important at t for him to

be right about whether or not P, and he does at t truly believe that P, but does so on the basis of evidence that, though probative to a degree, impresses us as too slight to confer on him knowledge that P. Then IRI seems to treat as on an equal footing either of two remedies: X can either improve his evidence; or he can work on his practical interests in such a way that much less is at stake whether he is right about P or not. He can grow his evidence to meet the standards for knowledge imposed by his practical interests at t; or he can so modify his practical interests as to shrink, as it were, the standards which knowledge that P requires. Suppose he takes the latter course. Then a situation may arise at a later time, t*, when we can truly affirm an 'ugly conjunction' like:

> X didn't (have enough evidence to) know P at t but does at t* and has exactly the same body of P-relevant evidence at t* as at t.

Such a remark seems drastically foreign to the concept of knowledge we actually have. It seems absurd to suppose that a thinker can acquire knowledge without further investigation simply because his practical interests happen so to change as to reduce the importance of the matter at hand. Another potential kind of ugly conjunction is the synchronic case for different subjects:

> X knows that P but Y does not, and X and Y have exactly the same body of P-relevant evidence.

when affirmed purely because X and Y have sufficiently different practical interests. IRI, as we noted earlier, must seemingly allow that instances of such a conjunction can be true.[12]

So far, so bad for IRI. But does contextualism escape any analogue of these problems? Certainly, there can be no commitment to either form of ugly conjunction so long as we are concerned with cases where the relevant standards are set as those of an ascriber distinct from X and Y. In that case the same verdict must be returned about X at t and at t*, or about X and Y, simply because some single set of standards is in play. But what if the context is one where contextualism has gone *flexible*, availing itself of the licence to defer to standards set by the (changing) pragmatic characteristics of the subject(s)? In that case, *non-indexical* contextualism, at least, can offer no evident barrier to the assertibility in suitable circumstances of either type of ugly conjunction. So much is simply

the price of the flexibility it appropriates to accommodate the cases that seemed to favour IRI.

Regular (indexical) flexible contextualism, by contrast, stands to suffer a commitment only to the metalinguistic counterparts:

> "X doesn't (have enough evidence to) know P" was true at t but "X does (have enough evidence to) know P" is true at t* and X has exactly the same body of P-relevant evidence at t* as at t;
>
> "X knows that P" and "Y does not know that P" are both true and X and Y have exactly the same body of P-relevant evidence.

These are spared 'ugliness' by the postulated shifts in the semantic values of the occurrences of "know" which are the trademark of the classical contextualist view and block disquotation. Nevertheless, they are unquestionably extremely strange to an English ear.

Does knowledge relativism fare better with these potential snags? Again, the interesting question concerns a flexible relativism: one with the resources to handle cases where the pragmatic features of its subject determine the standards that a correct knowledge-ascription has to meet. And of course for the relativist, as for the non-indexical contextualist, there are no complications occasioned by shifts in the semantic value of "knows". We know to expect that relativism will coincide in its predictions with non-indexical contextualism in all scenarios where knowledge is ascribed in the indicative mood and where there is no contrast between the ascriber and an assessor. It is therefore no more than the price paid for the flexibility to copy the verdicts of IRI in cases that reflect well on the latter that relativism, like non-indexical contextualism, will sanction certain cases, both synchronic and diachronic, of ugly conjunctions.

So here is the scorecard.

IRI is, seemingly, encumbered by a commitment to the assertibility, in suitable circumstances, of both forms of ugly conjunction.

However, commitments of this kind are not, as is sometimes assumed, a distinctive problem for that particular form of variabilism:

Non-indexical contextualism and *relativism* both share that commitment provided they avail themselves of the option of 'flexibility'. And of course, if they do not so avail themselves, the IRI-favourable cases stand as counterexamples to their proposals.

Classical (flexible) contextualism is committed only to metalinguistic versions of ugly conjunctions. That is not as bad only provided (i)

the metalinguistic versions are not as ugly and (ii) their disquotation is indeed blocked, i.e. provided "knows" is indeed context-sensitive.

7 Is There Any Good Reason to Think That "Knows" is Context-Sensitive?

When utterances of the same type-sentence in different contexts appear to be able to take differing truth-values, context-sensitivity—that is, sensitivity of the content expressed to features of the utterance-context—is plausibly the most natural explanation. So, anyway, it must have seemed to the original authors of contextualism when first reflecting on the apparent variability of "knows". But that was before the rival invariantist kinds of explanation here considered entered the scene. Can evidence be mustered to restore the presumption that context-sensitivity is at the root of the variability phenomena and so give classical contextualism an edge?

The literature on the matter is complex, extensive and inconclusive; it is fair to say that there are no uncontroversial, or even generally agreed criteria for (non-) context-sensitivity.[13] Jason Stanley argues persuasively[14] that the alleged context-sensitivity of "knows" is not felicitously assimilated to that of any of gradable adjectives ("rich", "tall"), pronouns ("I", "you", "this"), or quantificational determiners ("all", "many", "some"). Schaffer and Szabo grant this but suggest instead a comparison with so called A-quantifiers ("always", "somewhere").[15] Still, there is no reason in any case why a *bona fide* context-sensitive expression should behave exactly like context-sensitive expressions of other kinds. Can any *general* reason be given to think that "knows" and its cognates are context-sensitive, whether or not their behaviour sustains close comparison with that of other, uncontroversially context-sensitive expressions?

Here is a natural litmus. If "S" contains context-sensitives, then distinct tokens of "S" in different mouths may have different truth-conditions. So distinct token questions, "S?" in the mouths of different questioners may impose different conditions on the appropriateness of an affirmative answer. Hence if "knows" and its cognates are context-sensitive, it should be possible to design a pair of conversational contexts within which a pair of tokens of the question, "Does X know that P?" presented simultaneously to a single agent—the *questionee*—can respectively properly deserve prima facie conflicting answers.

Call this the *Forked Tongue test*. It's pretty crude—it won't, for instance, distinguish context-sensitivity from simple ambiguity. Still, its credentials as at least a necessary condition for context-sensitivity seem good. Let's construct a simple illustration. Suppose Ashley and Bobbie are wondering whether to duck out of the queues at the bank and go to get coffee and cake. Chris meanwhile, standing nearby, is on the phone to Denny. Bobbie overhears Chris say "Yes, my dear, there is. There is a Caffè Nero just two minutes away where they serve excellent coffee and *torta di cioccolata*." Bobbie says, "Excuse me, but did you say that there is a nice coffee shop just two minutes away." Chris replies, "Ah. Actually, no. I mean: I did say that, but I was talking to my partner about a location downtown."

Thus: "just two minutes away" passes the Forked Tongue test. It was the context of Denny's question, rather than Bobbie's, that set the reference of "just two minutes way" in Chris's original remark. When Bobbie puts a token of essentially the same type-question, the reference shifts and the correct answer changes.

Can we get a similar result with "know"? Let's try to construct an analogously shaped case, but where the questioners' respective contexts differ in respect of the stakes they have in the truth of the answer. So

Case 7: Ashley and Bobbie are dithering in the foyer of the bank as before. They talk about the risk of foreclosure and Bobbie says, "Look, we had better ask someone." Chris and Denny standing near the back of one of the queues, happen to overhear their conversation. Denny is also perturbed by the length of the queues and says to Chris, "Do you know if the bank will be open tomorrow? We could come back then if it will, but I'd rather not leave it till Monday since I have a hairdresser's appointment on Monday morning and am meeting Stacy for coffee in the afternoon." Chris, recalling the Saturday morning visit of two weeks earlier says, "It's OK. I happen to know the bank will be open tomorrow. I'll drive you over after breakfast." Ashley, overhearing, says, "Excuse me, but did you say that you know the bank will be open tomorrow?" Chris, mindful of Ashley and Bobbie's overheard priorities, replies, "Ah. Actually, no. I mean, I did use those words, but I was talking to Denny here, who has less at stake than you guys."

Case-hardened contextualists may find this dialogue unexceptionable, but I would suggest that Denny, Ashley and Bobbie might reasonably be baffled by Chris's last reply. It is also striking that, if the

dialogue *is* regarded as unexceptionable, it should remain so if all play with "know' is dropped and the operative question is rephrased as simply, "Will the bank be open tomorrow?" But in that case the explanation of the acceptability of Chris's final remark will presumably have nothing to do with context-sensitivity in the operative question. So it looks as though the contextualist faces a choice between admitting that "know" fails the Forked Tongue test in this instance, or insisting that it passes but that this fact has no significance for its putative context-sensitivity.

8 Is There Any Good Reason to Think that "Knows" is **Not** Context-Sensitive?

The consideration that has proved perhaps the most influential in this regard in the recent debates, and indeed has provided the prime motivation for knowledge relativism, is provided by ostensible patterns of *correction and retraction* that our knowledge-talk seems to exhibit. Here's a toy example of the relevant kind. Chris and Denny have gone away for the weekend and have left Ashley and Bobbie the keys for the use of their car.

> *Ashley*: Do you know where their car is parked?
> *Bobbie*: Yes, I do—Chris texted me that they left it in the multi-storey lot as usual after badminton on Friday.
> *Ashley*: But, as you very well know, there have been several car thefts in the neighbourhood recently. We should have gone to get it earlier. What if it's been stolen?
> *Bobbie*: I wasn't reckoning with that. OK, I guess I don't *know* that it is in the multi-storey lot—we had better go and check.

Here, the reader is intended to understand, Ashley's second question doesn't change Bobbie's epistemic situation—doesn't give her any more evidence. But it does persuade her that it is appropriate to impose more demanding standards of evidence on her answer than she started out doing—and she now disavows the knowledge she originally claimed.

Now, the crucial point for the relativist is the suggestion that this disavowal is to be understood as a *retraction*. Consider this continuation of the dialogue:

Ashley: Was your first answer, about knowing where the car is, true when you originally gave it, before I raised the possibility of the car's being stolen?

and two possible responses:

Bobbie: *Either* (a) Sure, but I could not truly repeat the words I used, once I was reminded of the recent incidence of car-theft.

Or (b) No; as I just said, I wasn't thinking about the possibility of the car's being stolen. I shouldn't have claimed to know that it is in the multi-storey lot.

The relativist's idea is that contextualism ought to predict that answer (a) can be acceptable. For if the content of a knowledge-ascription is relative to standards set by the context of ascription, then suitable changes in that context may be expected to go along with a shift in content consistent with tokens of a single type-ascription being respectively true in an original context but false in a later. But in fact answer (a) is, on the face of it, simply bizarre, and the natural answer, in context, is answer (b), which notably not merely supplants but critiques and retracts the original. That is evidence, it is alleged, that the content of the knowledge claim has not shifted in response to the change of standards, but has remained invariant throughout.

Note that the contextualist can of course allow Bobbie to affirm not merely that she doesn't know now where the car is but that she *didn't know* when she made her first answer. That is because the referent of "know", even as used in that past tense claim, will—according to contextualism—have shifted to some high-standards knowledge relation in response to Ashley's invoking the possibility of theft, whereas Bobbie's original claim will have involved some different, low-standards relation. So contextualism can actually predict what *sounds like* a retraction: "I didn't know that P". What, the critic will charge, it cannot predict is agents' willingness to treat such remarks *as* retractions—their refusal to stand by the different thing that, according to contextualism, they originally said.

9 But are the Retraction Data Solid? A Doubt

It is, however, a further question whether our patterns of apparent retraction of knowledge claims really *do* provide the powerful argument for relativism that its supporters, notably MacFarlane, have urged. I'll canvass two doubts.

To begin with, there are issues about what exactly should count as the manifestation in practice of the relevant kind of retraction. Do we, in response to changes in pragmatic factors, really retract former ascriptions of knowledge in exactly the sense that relativism needs? We have already noted an important distinction in this connection. Consider this dialogue:

> *Ashley (on a fast moving train in New Mexico)*: Look, there is a cougar!
> *Bobbie*: Where? I don't see it.
> *Ashley*: Just there, crouching by those rocks.
> *Bobbie*: I still don't see it.
> *Ashley*: Oh, I am sorry. I see now that it was just a cat-shaped shadow on the rocks. *There wasn't a cougar.*

Here Ashley's last speech is a retraction in anyone's book: she is denying, using appropriately changed context-sensitive language, exactly the thing she originally said. But to accomplish this, it suffices merely to change the tense of the original and negate it. Whereas under the aegis of classical contextualism about "knows", corresponding moves do *not* suffice for retraction of a knowledge-ascription, as we observed. Contextualism allows that Bobbie may perfectly properly admit, in response to Ashley's canvassing the possibility of car-theft, both that she does not know where the car is and *did not know when first asked*. The latter admission is not a retraction of the original claim, since—according to contextualism—it concerns a different, high-standards knowledge relation. Accordingly, the relativist needs to point to clear evidence in our linguistic practice that the disposition to retract knowledge claims when the stakes are raised goes deeper than the apparent denial involved in merely changing the tense and negating the result. Speakers will have to be reliably and regularly disposed to say things that distinguish what they are doing from such merely apparent retractions, since contextualism can take these in stride.

What kinds of sayings would manifest that distinction? Bobbie was presented above as doing something of the needed sort by saying "I wasn't thinking about the possibility of the car's being stolen. *I shouldn't have claimed to know that it is in the multi-storey lot.*" But that is exactly *not* what she should say on the assumption of knowledge relativism. Relativism allows that the earlier claim, in the lower-standards context then current, can have been perfectly appropriate—indeed, from the

standpoint of that context, *true*. So if that were the form that retractions of knowledge claims were generally to assume, the fact would be at odds with rather than advantageous to relativism. What is wanted, it seems, is a form of repudiation which is neither a simple denial, modulo any needed changes in tense, etc., nor a repudiation of the propriety of one's making the earlier claim in its original context.

The salient remaining possibility is something along the lines of, "What I said before is false". Unfortunately for relativism, even this pattern of retraction, should it be prevalent, is too coarse to be unpredictable by contextualism. The reason it is so is because in order to give what passes as an appropriate disquotational specification of what was said by some utterance in a previous context—"What he expressed before by S was that P"—it is not necessary—or indeed possible—to adjust *every* kind of context-sensitive expression that S may have contained. To be sure, if Ashley says, "Right now, I am going crazy waiting in this queue", then in order to specify what she said, we'll need to shift pronouns and tenses and temporal adverbs in routine ways: what Ashley said was that, *at that time*, *she was* going crazy waiting in *that* queue. But this does not apply in general to, for instance, gradable adjectives nor, so the contextualist may contend, to "knows" and its cognates. If an inexperienced hospital theatre orderly asserts, "This scalpel is very sharp", intending roughly that you could easily cut yourself if handling it carelessly, he may quite properly be reported to an expert surgeon as having said that *that particular scalpel is very sharp*, even when the context set by conversation with the surgeon is understood as one in which the notion of an instrument's sharpness is high-standards—for instance, is tied to its suitability for refined neurosurgery. And in such a context, the orderly may have to accept a reprimand and allow that "What I said—viz. that that scalpel is very sharp—was false." In short: where some kinds of context-sensitive language are involved, admissible ways of specifying 'what was said' are not guaranteed to deliver an actual content previously asserted rather than a counterpart spawned by differences between the original context of use and the context of the specification.

Of course it's usually easy enough to disambiguate in such cases if the conversational participants find it important to do so. The hospital orderly may (perhaps unwisely) protest that all he meant was that the scalpel had enough of a fine edge to be dangerous if handled carelessly. Perhaps therefore the relativist argument should be that we don't go in for such disambiguation where knowledge claims are concerned but, as

it were, *simply* retract. But is that true? With "sharp" now annexed to high (neurosurgical) standards, the orderly has to have recourse to other language to explain what he originally meant to say. If that is allowed to constitute sticking by his former claim, then we surely will want to say something similarly exculpatory about the credentials of our erstwhile epistemic situation and an associated knowledge claim even as we feel obliged to revoke the latter purely because of pressure of elevated standards.

It is, accordingly, open to question whether relativists have succeeded in tabling a notion of retraction with each of the needed features (a) that we do go in for retraction of knowledge claims under changes of pragmatic parameters of context, (b) that relativism predicts this and (c) that contextualism cannot predict as much.

10 But are the Retraction Data Solid? A Second Doubt

A second doubt about the alleged pro-relativistic significance that our patterns of retraction of knowledge claims supposedly carry concerns the *extent* of the phenomenon. Relativism predicts that two contexts of assessment, c_1 and c_2, differing only in the values of pragmatic parameters, may be such that one mandates an endorsement of a knowledge-ascription and another its repudiation. The examples so far considered have tended to focus on one direction: where a knowledge-ascription is made in a relatively low-standards context and then, apparently, retracted as the stakes rise, or certain error-possibilities become salient, or whatever the relevant kind of change is proposed to be. What about the converse direction? Does our practice pattern as relativism should expect?

Let's try an example:

Case 8 begins exactly as Case 2: It is Friday afternoon, and Ashley and Bobbie have arrived at the bank to deposit their salary cheques. However there are long queues at all the bank counters. Ashley recalls being at the bank on a Saturday morning two weeks ago and says, "Let's come back tomorrow. **I know the bank will be open tomorrow morning**." Suppose that the bank will indeed be open on the Saturday morning. However the couple's mortgage lender has written to say the company will foreclose unless the cheques are in the account by Monday to service

the monthly repayment, and Ashley and Bobbie are mindful of this. Bobbie says, "But what if the bank has changed its opening hours? Or what if the Saturday morning opening was some kind of one-off promotion?" Ashley says, "You're right. **I suppose I don't really *know* that the bank will be open tomorrow** (even though I am pretty confident that it will). We had better join the queue."
Invited intuition: Ashley correctly retracts her original claim. There is too much at stake to take the risk of e.g. a change in banking hours.

But now let's run the example on. Let it so happen that Eli, who is the manager of the local branch of Ashley's and Bobbie's mortgage company, is also waiting in one of the queues and overhears their conversation. Remembering 'that nice young couple' and taking pity on them, Eli comes across and says, "Don't worry, guys. Just between us, there is a degree of bluff about these 'final reminder' notices. We never actually foreclose without first making every effort to conduct an interview with the borrowers. It will be absolutely fine if this month's payment is serviced by the end of next week." Ashley and Bobbie are mightily relieved and Ashley says, "Aha. **So actually I *did* know that the bank will be open tomorrow**! Let's go and get a coffee and come back then."

Relativism predicts that Ashley's last emboldened remark is perfectly in order—indeed it expresses a commitment: the context after Eli's intervention is once again low-stakes, so low-standards, so Ashley's knowledge claim is now mandated by the original evidence and the intermediate knowledge denial should be retracted. But while relief and the decision to get a coffee are reasonable enough, Ashley's last remark is actually utterly bizarre.

This is a crucial issue for knowledge relativism. I have no space here to pursue it in detail, but I conjecture that there are actually no clear cases where, moving from a high- to a low-standards context, and *mindful of the fact*, we are content, without acquiring any further relevant evidence, simply to retract a former knowledge-disclaimer and to affirm its contradictory. Where P was the proposition of which knowledge was denied, we may well say things like, "Well, I guess it's reasonable now if we take it that P" or "We can now probably safely assume that P". But the claim to now *know* that P will simply invite the challenge to re-confront the error-possibilities made salient in the previous high-standards context. And when the changes involved in the context shift are wholly pragmatic, we will tend to regard ourselves as, strictly, no better placed,

epistemically, to discount those possibilities than we were before. For example, Ashley should not now after conversation with Eli, any more than earlier, want to claim *knowledge* that the Saturday opening of two weeks ago was not a one-off promotion.

The qualification, "mindful of the fact" is crucial. No doubt it may happen that, forgetting altogether about a previous high-standards situation, we may in a new, relaxed context be prepared to make knowledge claims that contradict earlier disclaimers. But these claims will properly rank as *retractions* only if we recall the previous context and what we said then. And if we do that, recollection of the error-possibilities that drove the early disclaimers is still likely to inhibit our outright claiming the relevant bits of knowledge even if it no longer seems urgent to reckon with those possibilities. Relativism, by contrast, predicts that there is now a mandate for such claims and that any such inhibitions about them conflict with the correct semantics for "knows".[16]

11 Conclusion

Variabilism, in all its stripes, is motivated by an *appearance*: that the language game of knowledge-ascriptions and denials incorporates a dependence of their truth-values on pragmatics—on interests, or saliences, or stakes. Each of the four theoretical proposals here considered, albeit offering very different accounts of the nature of the dependence involved, takes this appearance to be veridical. If, as has been the general tendency of the foregoing discussion, none of these accounts is satisfactory—if each under-predicts (fails to predict some uses) or over-predicts (predicts uses with which we are uncomfortable)—the natural conclusion is that the appearance is *not* veridical: that our discourse involving "knows" and its cognates is subject to no genuine pragmatics-sensitive variability of truth-conditions.

If we draw that conclusion, two possibilities remain. One, of course, is invariantism. But invariantism must come to a view about where the invariant threshold for knowledge falls, and wherever it is placed, it will have to be acknowledged that a significant body of our knowledge claims, or knowledge-disclaimers, are false and an explanation will therefore be owing of why so much of our linguistic practice with "knows" and its cognates falls into error. Invariantists have not been slow to respond to this challenge.[17] I here record the opinion, for which I have no space to argue, that to date their efforts have been unpromising.

The other possibility is a view concerning "knows" and its cognates that stands comparison with what deflationists about truth say about "true". For the deflationist about truth, very familiarly, it is a metaphysical mistake to ask after the character of the property that "true" expresses. The proper use of the word is accountable, rather, not to the nature of an assumed referent in the realm of properties but to the service of certain practical purposes—notably indirect endorsement, and generalisation—that it enables us to accomplish. Correspondingly, a deflationism about knowledge will discharge the idea that there is any determinate epistemic relation or—in deference to contextualism—family of relations that the proper use of "knows" serves to record and whose character determines the truth-conditions of knowledge-ascriptions. Rather the use of the word needs to be understood by reference to the practical purposes—notably, for example, as DeRose observed, the accreditation of potential informants and the appraisal of agents' rational performance—that it enables us to accomplish. The variability phenomena surface as one or another of these purposes comes to the fore in a particular pragmatic context. But it is a metaphysical mistake to project these phenomena onto the putative nature of an assumed referent, or referents, as IRI and contextualism attempt to do, and seek to explain them thereby.

Relativism doesn't make *that* mistake. Someone who holds that "X knows that P" is assessment-sensitive has already discharged the realism about the knowledge relation that deflationism would counsel us against. But if the suggestion of the preceding section about the asymmetries between our apparent retractions of knowledge-ascriptions and apparent retractions of knowledge-disclaimers are correct, then the concept of knowledge we actually have betrays an (inflationary) invariantist tendency which relativism simply misdescribes. Of course it is open to a relativist to acknowledge this, and to present relativism as reformist. That proposal, however, stands in need of an argument that any purpose would be served by reform. The essence of the case for deflationism about "knows" is two-fold: negatively, that the combination of our tendency to allow the standards for its application to inflate indefinitely while unwilling to accept, with the sceptic, that it never applies, betrays a concept with certain inbuilt tensions and no determinate reference; positively, that the word nevertheless supplies the valuable resources that the variability phenomena reflect. The first part of that might suggest the desirability of reform. But that is compensated for by the second.

Such a general conception of knowledge—or better: of the function of "knows"—is nothing new,[18] although the present suggestion, that its correctness is the principal lesson which the variability phenomena have to teach us, may be so. If it is correct, there is a striking corollary: the idea that knowledge should come first in analytical epistemology could not be more misguided. Rather, knowledge—the putative substantive referent of "knows"—comes nowhere.

New York and Stirling Universities

Notes

1. I draw substantially on my 2017 "The Variability of 'Knows': An Opinionated Overview" in Jonathan Jenkins Ichikawa (ed.) *The Routledge Handbook of Epistemic Contextualism*, Routledge. pp. 13–31.
2. This second aspect—Williamson's 'Knowledge First' programme—is of course strictly independent of and additional to the primitivism.
3. Cohen (1986), DeRose (1992).
4. Hawthorne (2004), Stanley (2005); see also Fantl and McGrath (2007).
5. Or sometimes: subject-sensitive invariantism.
6. MacFarlane (2005, 2014).
7. As MacFarlane terms it.
8. For experimentation with a version of this kind of view, see Cappelen (2008). Weatherson (2009) makes an interesting application of it to address certain puzzles with indicative conditionals.
9. This is different to—but perhaps not quite as clean cut as—Stanley's (2005) tactic which is to develop examples where a low-stakes ascriber does not know that the subject is high-stakes. For instance, suppose Chris does not notice Ashley and Bobbie join the queue. But Denny, Chris's partner, who has overheard the exchange, does and nudges Chris with a quizzical glance in their direction. Chris says: "Oh, I guess they must have remembered some reason why they can't come back tomorrow—after all, **they now know that the bank will be open then**." This time, we are supposed to have the intuition that the knowledge-ascription is false.
10. See DeRose (2009, chapter 7).
11. John Hawthorne (2004, chapter 4 at pages 162–166) attempts to enlist the help of what he calls the "psychological literature on heuristics and biases". Hawthorne's idea is that one lesson of this literature is that the becoming salient of a certain risk in a high-stakes situation (e.g. that of the bank's changing its opening hours) characteristically leads us to overestimate its probability in general and hence to project our own

ignorance onto subjects in low-stakes situations too. DeRose (2009, chapter 7, section 3) counters that the phenomenon to be explained—high-stakes agents' denial of knowledge to low-stakes subjects—extends to cases where the former take it that they *do* nevertheless know the proposition in question (because they take themselves to meet the elevated standards demanded by their high-stakes context). That seems right, but I do not see that Hawthorne needed the "projection of ignorance" component in his proposal in any case; a tendency to overestimation of the probabilities of salient sources of error would seem sufficient to do the work he wants on its own. The objection remains, however, that if an overestimation of the risk of a certain source of error underlies a high-stakes ascriber's denial of knowledge to themselves, the good standing of that denial is already compromised—whereas IRI requires precisely that the high-stakes context should validate it.

12. If evidence, too, were an interest-relative notion, then a possible direction of defence for IRI against these ugly-conjunctive commitments would be to try to make a case that variation in the interests of a subject sufficient to make the difference between her knowing that P and failing to do so must also affect what evidence she possesses, thus undercutting the assumption that evidence may remain constant for a subject at different times, or for distinct subjects when their interests differ. Stanley canvasses this suggestion (2005, p. 181). It misses the nub of the difficulty, however, since there will presumably be cases where the relevant evidence is known with certainty and hence must be reckoned to be in common no matter what the practical interests of the subjects, or subject at different times.
13. For discussion, see Cappelen and Hawthorne (2009, chapter 2).
14. Stanley (2005, chapter 3).
15. Schaffer and Szabo (2013). Their proposal deserves a properly detailed discussion. I believe the comparison is flawed but I have no space to enlarge on that here.
16. This objection should be contrasted with another made by Montminy (2009). His contention is that when *in a high-standards context* we disclaim knowledge that P, we will also judge that we will be wrong to reclaim knowledge that P in a subsequent low-standards context, even though—he allows—that is what we will do once such a context is entered into and relativism says we will be right to do so. I agree with the first part of that—namely, that we will take a dim view, while in the high-standards context, of the envisaged subsequent reclamation and that since relativism says that there is nothing wrong with the subsequent reclamation, there is here a tension between something we are inclined to think and what relativism thinks we ought to think. But, unless I misread

Montminy, I'm saying something different and stronger as well: namely that we *won't actually make a retraction* of the previous knowledge denial when we get into the low-standards context.

Knowledge relativism, in other words, mispredicts not just aspects of our attitudes to our practice with "knows" but our practice itself. (MacFarlane responds to Montminy in section 8.6 of his (2014), see especially p. 198 and following. His response does not engage the objection made here.)

17. See, for example, Williamson (2005).
18. The germ is famously present in Austin (1946, pp. 97–103) where a view is outlined on which utterances of the form "I know that such-and-such" serve a *performative* rather than a descriptive function, and the function of "I know" is in effect to offer a *promise* of truth, on the basis of which others are entitled to act, form beliefs, or claim to know in turn. Austin's ideas receive a thoroughgoing, sympathetic development in Lawlor (2013), though I do not know how far she would welcome the deflationism prefigured here.

References

Austin, J.L. 1946. Other Minds. *Proceedings of the Aristotelian Society*, Supplementary Volume 20: 148–187.

Cappelen, H. 2008. Content Relativism and Semantic Blindness. In *Relative Truth*, ed. Manuel Carcia-Carpintero and Max Koelbel, 265–286. Oxford: Oxford University Press.

Cappelen, H., and J. Hawthorne. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Cohen, S. 1986. Knowledge and Context. *The Journal of Philosophy* 83: 574–583.

DeRose, K. 1992. Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research* 52: 913–929.

DeRose, K. 2009. *The Case for Contextualism: Knowledge, Skepticism, and Context: Vol. 1*. Oxford: The Clarendon Press.

Fantl, J., and M. Mcgrath. 2007. Knowledge and the Purely Epistemic: In Favor of Pragmatic Encroachment. *Philosophy and Phenomenological Research* 75 (3): 558–589.

Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

Lawlor, K. 2013. *Assurance: An Austinian View of Knowledge and Knowledge Claims*. Oxford: Oxford University Press.

MacFarlane, J. 2005. The Assessment Sensitivity of Knowledge Attributions. *Oxford Studies in Epistemology* 1: 197–233.

MacFarlane, J. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.

Montminy, M. 2009. Contextualism, Relativism and Ordinary Speakers' Judgments. *Philosophical Studies* 143 (3): 341–356.

Schaffer, J., and Z. Szabo. 2013. Epistemic Comparativism: A Contextualist Semantics for Knowledge Ascriptions. *Philosophical Studies* 2: 1–53.

Stanley, Jason. 2005. *Knowledge and Practical Interests.* Oxford: Oxford University Press.

Weatherson, B. 2009. Conditionals and Indexical Relativism. *Synthese* 166: 333–357.

Williamson, T. 2005. Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowledge. *The Philosophical Quarterly* 55 (2019): 213–235.

Wright, C. 2017. The Variability of 'Knows': An Opinionated Overview. In *The Routledge Handbook of Epistemic Contextualism* ed. Jonathan Jenkins Ichikawa, 13–31. Abingdon: Routledge.

Truth Relativism and Evans' Challenge

Sebastiano Moruzzi

1 Evans' Challenge

Gareth Evans (Evans 1979) has famously advanced an argument against the intelligibility of temporalism. The basic thought is that temporalism cannot use the analogy between the relativity of truth to worlds and the relativity of truth to times to make sense of the correctness conditions of the speech act of assertion when temporally neutral propositions are involved.

Evans' basic worry is simple: if the truth of a proposition a is relative matter, then what should I assert? Of course, much depends on what relativity we are talking about. Take the least controversial case of truth relativity: relativity to possible worlds.[1] Though propositional truth is relative to possible worlds, there is a straight answer to what I should assert: actual truths, actuality works as a criterion for privileging one world (our world) over the others. The answer to Evans' question becomes less straightforward when we consider another relativity thesis: the relativity of propositional truth to times, i.e. temporalism. Evans issues a challenge against this view in the following terms:

S. Moruzzi (✉)
Department of Philosophy and Communication Studies,
University of Bologna, Bologna, Italy
e-mail: sebastiano.moruzzi@unibo.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_18

> Such a conception of assertion is not coherent. In the first place, I do not understand the use of the ordinary word 'correct' to apply to one and the same historical act at some times and not at others, according to the state of the weather. Just as we use the terms 'good' and 'bad', 'obligatory' and 'permitted' to make an assessment, once and for all, of non-linguistic actions, so we use the term 'correct' to make a once-and-for-all assessment of speech-acts. Secondly …If a theory of sense permits a subject to deduce that a particular utterance will now be correct, but later will be incorrect, it cannot assist the subject in deciding what to say, nor in interpreting the remarks of others. What should he aim at, or take others to be aiming at? Maximum correctness? But of course, if he knew the answer to this question, it would necessarily generate a once-and-for-all. (Evans 1979: 349–50)

Evans' point can be extended to other forms of truth relativism that allow for a variation of propositional truth in the same world. Following Greenough (2011), I reconstruct Evans' argument against truth relativism in the following way[2]:

1. The question 'What should he aim at' when we make assertions is a legitimate question.
2. Any legitimate question has a legitimate answer.
3. Any legitimate answer to this question will generate a once-for-all answer.
4. Any once-for-all answer is incompatible with Truth Relativism.
5. Therefore, Truth Relativism is ruled out.

Contemporary truth-relativists have mainly objected to the latter argument by rejecting step 4: truth is relative but there is always, as in the case of world relativization, a criterion for providing a once-for-all answer to what to believe and assert.[3] These forms of truth relativism share the view that the norm for assertion and belief should be relativized to the parameter determined by the context of the asserter.[4]

2 Assessment-Sensitive Relativism, Assertion and Retraction

In this chapter I will consider MacFarlane's assessment-sensitive relativism. Assessment-sensitive relativism is the thesis that the 'relativity of truth –either of sentences or of propositions– to possible contexts of

assessment' is not an absolute matter. According to assessment-sensitive relativism, there are expressions whose extension is relative not only to the context of use but also to what MacFarlane calls *context of assessment*: a context with respect to which a past utterance is assessed. According to assessment-sensitive relativism, it is possible that according to one context of assessment a past assertion expressed a true proposition, whereas according to another context the same proposition expressed by the same speech act is false. As we will see later, this possibility allows the assessment-sensitive relativist to claim that it is possible to correctly criticise one's own past assertion by retracting it, though such an assertion was correct as assessed at the time of the utterance.[5] Such propositions are called *assessment-sensitive propositions*.

Crucially, the notion of context of assessment arises not only in the semantics, but also in the norms for speech acts. In MacFarlane's framework, there is one norm for assertion and one for retraction.

According to the latest version of MacFarlane's truth relativism, assertion is characterised by the following norm (MacFarlane 2014: 103)[6]:

Reflexive Truth Rule An agent is permitted to assert that p in context $\boldsymbol{c}_1$ only if p is true as used in $\boldsymbol{c}_1$ and assessed from $\boldsymbol{c}_1$.

The rule provides a necessary condition for assertion. Though truth is not a sufficient condition for having a permission to assert a proposition, untruth is a sufficient condition for a prohibition to assert a proposition. So, on MacFarlane's view, Evans' challenge can be met without abandoning a genuine form of truth relativism since the Reflexive Truth Rule gives a once-for-all answer to the question of when it is not permitted to assert a sentence.

The Reflexive Truth Rule (henceforth 'reflexive rule') is not sufficient for distinguishing assessment-sensitive relativism from its more moderate cousin: non-indexical contextualism, according to which a proposition can be true relative to one context of use and false relative to another. Without assigning a distinctive role to the contexts of assessment in the post-semantics, assessment-sensitive relativism and non-indexical relativism are indistinguishable as far as the reflexive rule for assertion is concerned. According to MacFarlane, an assessment sensitive relativist has thus to add a further norm related to the speech act of retraction MacFarlane (2014: 108):

Retraction Rule An agent in context $\boldsymbol{c}_2$ is required to retract an (unretracted) assertion of $\boldsymbol{p}$ made in $\boldsymbol{c}_1$ if $\boldsymbol{p}$ is not true as used in $\boldsymbol{c}_1$ and assessed from $\boldsymbol{c}_2$.

In English, the act of retraction is signalled by the expression 'I was wrong' (MacFarlane 2014: 13–14).

Consider a toy example that we will use for the sake of exposition. One day, Ji Sung, a young college student casually listening to a morning radio program, learns what Fermat's Last Theorem is.[7] Ji Sung still does not know anything about Wiles' proof and is sceptical about the soundness of the conjecture. He thus asserts in the morning (context c_m) that Fermat's Last Theorem might be false—call this proposition *FERMAT*. Later in the day, Ji Sung comes to know that Wiles published a proof. In the evening, chatting with some friends, (context c_e) he asserts 'I was wrong. Fermat's Last Theorem must be true since Andrew Wiles proved it.'

According to MacFarlane, these uses of 'I was wrong' in relation to the epistemic modals—i.e. 'might' and its dual 'must'—are linguistic evidence for the assessment sensitivity of epistemic modal expressions.

3 Relativistic Semantics and Post-semantics Clauses

More precisely, the assessment-sensitivity thesis of 'might' and 'must' amounts to the thesis that the extension of these epistemic modals is sensitive to the context of assessment. So, assuming that these expressions behave as operators, the semantics is given by the following clauses for the truth in a context and an index—consisting of a world, time, information state, and assignment—:

'might p' is true in context c and index (w,t,i,a) if 'p' is true in context c and index (w',t,i,a) for **some** w', where w' belongs to the set of the worlds constituting the information state i; otherwise 'might p' is false in context c and index (w,t,i,a).

'must p' is true in context c and index (w,t,i,a) if 'p' is true in context c and index (w',t,i,a) for **all** w' that belong to the set of the worlds constituting the information state i; otherwise 'must p' is false in context c and index (w,t,i,a).

The semantic clauses are neutral with respect to the assessment sensitive thesis. The assessment sensitive thesis kicks in in the post-semantics, which defines the truth normative notions profile of the speech acts of a language. Assessment-sensitive relativists employ, in their post-semantics, a doubly relativized truth predicate. In fact, the relativist's truth predicate is relativized to the context of use and the context of assessment of a sentence. On a relativist post-semantics, 'might be' is true in a context of

use c_1 in relation to the information available to a context of assessment c_2 (which might coincide with c_1):

'Might p' is true as used in c_1 and assessed from c_2 if 'Might p' is true in context c_1 and the index (w, t_{c_1}, i_{c2}, a) where t_{c1} is the time of the context of use and i_{c_2} is the information state of the context of assessment; otherwise 'Might p' is false as used in c_1 and assessed from c_2.

'Must p' is true as used in c_1 and assessed from c_2 if 'Must p' is true in context c_1 and at the index (w, t_{c_1}, i_{c2}, a) where t_{c1} is the time of the context of use and i_{c2} is the information state of the context of assessment; otherwise 'Must p' is false as used in c_1 and assessed from c_2.

If we adopt the above semantic and post-semantic clauses for Ji Sung's assertion, we have that Ji Sung's assertion of FERMAT (which is a might proposition) can be permitted in the morning since FERMAT is true as used in c_m and as assessed from c_m—the information available in c_m does not rule out the falsity of the theorem. In contrast, in the evening Ji Sung should retract his morning assertion since FERMAT is false as used in c_e and as assessed from the evening context c_e—the information available in c_e does rule out the falsity of the theorem.

According to MacFarlane, the intuitions of competent speakers of English confirm that retractions such the ones illustrated by the Ji Sung toy story are correct uses of epistemic modals. Thus the relativist post-semantics best captures the correctness conditions of the epistemic modal sentences in English. Or so MacFarlane argues.

In the rest of this chapter I will bracket the empirical adequacy of assessment-sensitive semantics though recent empirical data (Knobe and Yalcin 2014) cast some doubt on it in relation to epistemic modals. Rather, I will object to assessment sensitivity by producing an a priori argument which brings back Evans' challenge alive.[8] The argument is based on two assumptions: (1) that it is legitimate to translate the relativist post-semantics into the language of higher-order truth assessments; and (2) that it is legitimate to interpret the retraction data on the basis of these expressive resources. If we grant these assumptions, assessment-sensitive relativism is vulnerable to Evans' challenge.

What follows sketches the plan for the rest of this chapter. In Sect. 4, I introduce the language of higher-order truth assessment and provide a partial translation of it in relation to the truth predicate of the relativistic post-semantics. In Sect. 5, I present some principles of higher-order

truth assessment by means of the trumping framework and propose a way to express the retraction data by means of a trumping principle. In Sect. 6, I argue that by translating the reflexive rule into the terms of higher-order truth assessments and by using the trumping framework, a case for the relative permissibility of assertion can be made. In Sect. 7, I reply to an objection against the use of the trumping framework for interpreting the retraction data. In Sect. 8, I argue that the relative permissibility of assertion builds a case for a new version of Evans' challenge to assessment-sensitive relativism. Lastly, in Sect. 9, I deal with other objections.

4 Higher-Order Truth Assessments

Before proceeding, I need to translate the post-semantics vocabulary into a more compact vocabulary that I will use for presenting some thesis of higher-order relativism. This vocabulary is the vocabulary of higher-order truth assessments.

Let me introduce a two-place assessment truth predicate 'T(p, c)' to express the truth of the proposition *p* relative to the context of assessment *c*. This truth predicate can be embedded, giving rise to higher-order truth assessments:

First-order assessment $T(p, c_1)$
Second-order assessment $T(T(p, c_1), c_2)$
Third-order assessment $T(T(T(p, c_1), c_2), c_3)$
…and so on

This assessment predicate has no equivalent in MacFarlane's metalanguage, for it does not include the relativization to the context of use. As we have seen in Sect. 3, semantic and post-semantic clauses need a context of use for gathering the information necessary to initialise the index. So the notion of being true relatively to a context of assessment has in general no equivalent in the assessment-sensitive meta-language. However, as a special case, we can provide such a translation when we consider the truth assessment of a proposition expressing the act of assertion of a proposition *p* used in a context $c_1 - p_{c_1}$. Call this type of first-order assessment a *first-order assertion assessment*:

First-order assertion assessment $\boldsymbol{T(p_{c1}, c_2)}$

This first-order assertion assessment is translatable into the assessment-sensitive post-semantics in the following way:

$\boldsymbol{T}(\boldsymbol{p}_{\mathbf{c}1}, \boldsymbol{c}_2) =_{\text{def}} \boldsymbol{p}$ is true as used in $\boldsymbol{c}_1$ and assessed from $\boldsymbol{c}_2$.

A *second-order assertion assessment* is the assessment from a context $\boldsymbol{c}_3$ of a first-order assertion assessment $\boldsymbol{T}(\boldsymbol{p}_{\mathbf{c}1}, \boldsymbol{c}_2)$:

Second-order assertion assessment $\boldsymbol{T}(\boldsymbol{T}(\boldsymbol{p}_{\mathbf{c}1}, \boldsymbol{c}_2), \boldsymbol{c}_3)$

Strictly speaking, this second-order assessment has no direct translation into the two-place meta-linguistic truth predicate of MacFarlane's post semantics since we lack the context of use of the proposition $\boldsymbol{T}(\boldsymbol{p}_{\mathbf{c}1}, \boldsymbol{c}_2)$. However, note that $\boldsymbol{T}(\boldsymbol{p}_{\mathbf{c}1}, \boldsymbol{c}_2)$ is a meta-linguistic proposition stating some fact about the truth conditions of an assertion of a sentence at a certain context of use. Presumably, the facts relevant for the expression of the proposition and the index needed for the evaluation of this meta-linguistic proposition do not require any further information other than the information provided by the context of use $\boldsymbol{c}_1$. The reason for this is twofold: first, assuming that the meta-linguistic expressions are not indexicals, all the other indexical expressions will receive a value from $\boldsymbol{c}_1$; second, assuming that there are no elements of the index that are characteristic of the metalanguage vocabulary, the needed elements of the index, such as world, time, information, standards of taste, etc. … are already determined by $\boldsymbol{c}_1$ or $\boldsymbol{c}_2$ when needed. In other words, I put forward the *invisibility conjecture*: the context of use of an assertion assessment is semantically invisible—it does not introduce any element relevant for the indexing of the meta-linguistic proposition that is not already provided by the context of use of the assertion. The invisibility conjecture is no new phenomenon, in fact we have a similar case with

(K) the assertion of <The King of France is bald> made in the actual world at the time of Louis XIV is true relatively to the actual world.

In order to assess the truth of the proposition expressed by *K* the index involving the world, time, information, etc., of the context of use of *K* has no role in initialising the index of *K*: if *K* is uttered in a world of hairy kings, such a world has no role for the evaluation of *K*.[9]

In other words, if we name $c_{LouisXIV}$ the context of use of <The King of France is bald> in relation to the assertion of *K* made in the actual world at the time of Louis XIV, we have the following equivalence:

K is true as used in c_1 and as assessed from c_2 if *K* is true as used in $c_{LouisXIV}$ and as assessed from c_2.

Crucially, note that the invisibility conjecture does not rule out that the context of assessment could be relevant for the initialisation of the index of evaluation, it only excludes that the context playing the role of

the context of use of an assertion assessment has any relevance for the initialisation of the index.

If the invisibility conjecture is correct, then we can provide a translation of a second-order assertion assessment into the assessment-sensitive post-semantics as follows:

$\boldsymbol{T}(\boldsymbol{T}(\boldsymbol{p}_{\boldsymbol{c}_1}, \boldsymbol{c}_2), \boldsymbol{c}_3) \ = \ _{\mathbf{def}}[\boldsymbol{p}$ is true as used in $\boldsymbol{c}_1$ and assessed from $\boldsymbol{c}_2]$ is true as used in $\boldsymbol{c}_1$ and as assessed from $\boldsymbol{c}_3$.

In other words, the context of use can be compressed so as to show up only in the subscript of the assertion.

This translation procedure is generalisable, since what counts as the context of an assertion for the translation of any *n*th order assertion assessment is the context of use mentioned in the nested proposition of the assessment:

$\boldsymbol{T}(\boldsymbol{T}(\boldsymbol{T}(\boldsymbol{p}_{\boldsymbol{c}_1}, \boldsymbol{c}_2), \boldsymbol{c}_3) \ldots \boldsymbol{c}_{\boldsymbol{n}}) \ = \ _{\boldsymbol{def}}[[\boldsymbol{p}$ is true as used in $\boldsymbol{c}_1$ and assessed from $\boldsymbol{c}_2]$ is true as used in $\boldsymbol{c}_1$ and assessed from $\boldsymbol{c}_3]$ … is true as used in $\boldsymbol{c}_1$ and as assessed from $\boldsymbol{c}_{\boldsymbol{n}-1}]$ is true as used in $\boldsymbol{c}_1$ and as assessed from $\boldsymbol{c}_{\boldsymbol{n}}$.

In conclusion, even if I cannot give a general translation schema of the higher-order truth assessment predicate to the truth predicate of the relativistic post-semantics, I can provide such a translation for any *n*th order *assertion* truth assessment.

5 Trumping and Retraction

In the previous section I introduced the language of truth assessment and provided a (partial) translation into the language of relativistic post-semantics. In the present section I will introduce some principles relating higher-order truth assessments—I will refer to this set of principles as *the trumping framework*—and I will connect these principles to retraction.

To introduce the trumping framework, I will first explain the notion of trumping and then introduce some principles governing trumping.

Let's say that the assessment context c_1 *trumps* another one c_2 when the facts as assessed relative to the assessment context c_1 determine, from that context, their proper assessment relative to c_2. Among the possibilities for such a dependency, Moruzzi and Wright (2009) draw attention to two cases. First, we have *inward trumping* with respect to an assertion p_c just in the case that whenever p_c is true relative to a context of

assessment, it is thereby true from that same context of assessment that p_c is true relative to *any* other context of assessment. That is:

Inward Trumping $\forall c_x, c_y, c_z\,[\boldsymbol{T}(p_{c_z}, c_x) \rightarrow \boldsymbol{T}(\boldsymbol{T}(p_{c_z}, c_y), c_x)]$

Second, we have *outward trumping* when the determination is in the other direction:

Outward Trumping $\forall c_x, c_y, c_z\,[\boldsymbol{T}(\boldsymbol{T}(p_{c_z}, c_y), c_x) \rightarrow \boldsymbol{T}(p_{c_z}, c_x)]$

I will now use the trumping framework to express assessment-sensitive relativism.

Let's go back to the example of Ji Sung. It is natural to think that the admission of being wrong however cannot but be the admission that one context of assessment trumps another: it is true from the evening perspective (context c_e) that the morning assertion is untrue relative to the morning perspective. In other words, when Ji Sung occupies the later context c_2 he can correctly assess as false her morning assertion of p by using the 'I was wrong' expression.

I follow Moruzzi and Wright (2009: 321–323) in interpreting Ji Sung's use of 'I was wrong' as excluding the possibility of qualifying it with 'though it was correct in the context which I said that p'. Such use of 'I was wrong' is naturally read as an act of trumping.

The idea, then, is that when in the evening Ji Sung correctly (according to MacFarlane) says that his morning assertion was wrong, his retraction is tantamount to an act of trumping: his evening assessment context determines that, relatively to the evening assessment context, the morning assertion of *FERMAT* was false relatively to the evening assessment context and thus that her assessment in the morning of *FERMAT* as true was wrong. If we call $\neg FERMAT_{c_m}$ the assertion of $\neg FERMAT$ in the morning context c_m, and c_e is the evening context, we can express the latter trumping thought as the following instance of inward trumping:

$$T(\neg FERMAT_{c_m}, c_e) \rightarrow T(F(FERMAT_{c_m}, c_m), c_e) \qquad (1)$$

(1) says that it if the assertion of $\neg FERMAT$ in the morning is true from the evening perspective, then the evening assessment contexts forces the assessment as false of the assertion of *FERMAT* in the morning.

Let me be clear about this claim. I am aware that MacFarlane has never used something similar to (1) to express retraction. However, what I claim is that the language of higher-order assessments can be used to describe that linguistic data of retraction by means of (1).[10]

6 Higher-Order Assessments and the Reflexive Rule

Before proceeding I need to make clear what is the interaction between the trumping framework and the reflexive rule for assertion. The reflexive rule for assertion says that an agent is permitted to assert a proposition in a context only if the proposition is true as used *and assessed* from that context. Thus, in our schematic notation, the reflexive rule dictates that it is permitted to make the assertion p_{c_m} only if $T(p_{c_m}, c_m)$.

An instance of the reflexive rule in relation to our example is thus the rule according to which it is permitted to assert *FERMAT* in the morning only if its assertion in the morning is true as assessed from the very same morning perspective:

Reflexive Rule FERMAT (RRF) It is permitted to assert ***FERMAT*** in $\boldsymbol{c_m}$ only if ***FERMAT*** is true as used in $\boldsymbol{c_m}$ and as assessed from $\boldsymbol{c_m}$.

Given the translation scheme defined in Section 4, this rule can be translated as follows:

Reflexive Rule FERMAT* (RRF*) It is permitted to assert ***FERMAT*** in $\boldsymbol{c_m}$ only if $\boldsymbol{T(FERMAT_{c_m}, c_m)}$.

The trumping framework introduces among the candidate propositions for being asserted also truth assessments of assertions. So the reflexive rule will have normative consequences for these propositions as well as for the more 'ordinary' propositions. What is relevant here is the complex proposition expressed by the consequent of 2:

$$T(T(\neg FERMAT_{c_m}, c_m), c_e) \tag{2}$$

Given the reflexive rule for assertion RRF*, (2) is tantamount to the admission that the assertion of *FERMAT* was not permitted in the morning from the perspective of the evening[11]: Once immersed in the evening perspective (c_e), it is false to hold that *FERMAT* is true relatively to the morning perspective (c_m). Hence, once immersed in the evening perspective, the assertion of *FERMAT* in the morning was not permitted. So from the evening perspective, the reflexive rule was violated in the morning.

The first conclusion I draw is thus that the reflexive rule can be violated even if the asserted proposition is true from the assessment context of the asserter call this first result *Relative Assertibility*. Crucially, Relative Assertibility neutralises the assessment-sensitive relativist's objection to step 4 of Evans' argument: it not true that a once-for-all answer is compatible with Truth Relativism: in fact, the permissibility of an assertion turns out to be a relative matter.

7 Two Ways of Being Wrong

The key idea behind the argument for Relative Assertibility is the use of inward trumping for expressing the retraction data. Against this idea, it might be objected that the use of inward trumping for expressing retraction is not faithful to what is expressed by 'I was wrong'. The point of the objection is that the use of inward trumping seems to commit one to the admission of a fault, contrary to what MacFarlane intends. In fact, MacFarlane (2014: 240, footnote 3) distinguishes two senses of saying 'I was wrong':

1. saying that one was wrong *in claiming* p;
2. saying that one was wrong *to claim that* p.

According to MacFarlane, only the latter reading of 'I was wrong' is tantamount to an admission that the speaker should not have made the assertion in the first place. But if that were the case, then this notion of fault would have consequences for the correctness of the past assertion.

The point of this objection is thus that the inward trumping reading of retraction is not faithful to what is expressed by 'I was wrong' since it relies on the second reading: only the second reading expresses the admission the speaker should not have made the assertion, whereas MacFarlane intends the first reading in relation to retraction data. With the first reading, it is possible to say that the morning assertion of *FERMAT* was not faulty (and hence permitted) *consistently* with retraction data. At the same time, the retraction data is expressed by means of the second reading.

MacFarlane does not elaborate more on this distinction. Whether or not the distinction is sound, I will show that the trumping framework allows expressing the retraction data consistently with the idea that the speaker was not wrong in making the assertion.

I take it that with the first reading, MacFarlane believes that it is possible to say that the morning assertion of FERMAT was not faulty (and hence permitted) consistently with the retraction data. I submit that we can express this condition by means of the thought that whereas the morning perspective assesses as true the asserted proposition, the evening perspective assesses it as false:

I was wrong (IWW) $\boldsymbol{T(T(FERMAT_{c_m}, c_m), c_m) \wedge T(F(FERMAT_{c_m}, c_m), c_e)}$

Is IWW justified? I will now argue that both conjuncts of IWW can derived from the hypothesis that *FERMAT* is assessment sensitive and

that inward trumping holds. So the objection fails in holding that the trumping reading of the retraction data trades on a wrong reading of the data—i.e. a reading that is unsuitable to express the retraction data.

The first conjunct of IWW says that from the morning perspective (i.e. as assessed from c_m) it is true that that her morning assertion was true as assessed from the morning perspective (i.e. c_m again):

$$T(T(FERMAT_{c_m}, c_m), c_m) \tag{3}$$

In order to derive (3), note that Ji Sung's assertion is true from his morning perspective:

$$T(FERMAT_{c_m}, c_m) \tag{4}$$

Given that the following and almost trivial instance of inward trumping holds

$$T(FERMAT_{c_m}, c_m) \rightarrow T(T(FERMAT_{c_m}, c_m), c_m) \tag{5}$$

it follows that (5), together with (4), implies (3).

The second conjunct of IWW can be derived as follows. By hypothesis, what Ji Sung asserted in the morning is false relative to the evening context:

$$T(\neg FERMAT_{c_m}, c_e) \tag{6}$$

If we grant the inward trumping principle (1), it follows that its consequent is true:

$$T(F(FERMAT_{c_m}, c_m), c_e) \tag{7}$$

which is the second conjunct of IWW.

Note that because of the reflexive rule for assertion RRF*, (3) is tantamount to the admission that the assertion of *FERMAT* would be permitted in the morning from the morning perspective: once immersed to the morning perspective (c_m), it is true to hold that *FERMAT* is true relatively to the morning perspective (c_m), and hence, once immersed in the morning perspective, the assertion $FERMAT_{c_m}$ *would be* permitted as far as the truth value of *FERMAT* is concerned.[12]

However, (7) says that from the evening perspective it is true that ¬*FERMAT* asserted in the morning was true as assessed from the morning perspective. So, given (7) and RRF*, it follows that from the evening perspective Ji Sung's morning assertion of *FERMAT* *was not* permitted.

In conclusion, it is a *relative matter* whether Ji Sung was wrong in claiming *FERMAT*. Relative Assertibility is thus vindicated.

8 Back Again to Evans' Challenge

Let's recap what has been achieved so far. When Ji Sung occupies the context of assessment of the morning, he is committed to admitting that (4) holds. The reflexive rule for assertion, together with (3), justify the idea that from the morning perspective the morning assertion of *FERMAT* is permitted. On the other hand, (6), together with (1), imply that from Ji Sung's evening perspective his morning assertion is not permitted. So we have that:

i. assessment-sensitivity[13] together with the relevant inward-trumping principles[14] entail (7) and (3);
ii. (7) and (3) give rise to Relative Assertibility: the permissibility of asserting ***FERMAT*** is a relative matter.

In other words, assessment sensitivity, the inward-trumping principles (7) and (3) and the assertion rule RRF*, implies that the permissibility of *FERMAT* is a relative matter: according the context of assessment of the morning, the morning assertion of *FERMAT* would be permitted; whereas according to the context of assessment of the evening, the morning assertion ***FERMAT*** could not be permitted. Hence Relative Assertibility is vindicated.

Should then Ji Sung have asserted *FERMAT* in the morning or not? We are back to Evans' challenge! Instead of framing the challenge in terms of what is correct to assert, the challenge is framed in terms of what is permitted to assert. The assessment-sensitive relativist tried to block the challenge by adopting the reflexive rule for assertion. However, the considerations related to retraction and trumping show that this answer fails to secure the thesis that there is a once-for-all answer for the permission of an assertion that is compatible with assessment-sensitive relativism. But if no once-for-all answer for the permission of an assertion is available, then Evans' challenge is still open.

9 Objections and Conclusion

Before finishing, let me address two objections.

Objection 1

The argument for Relative Assertibility employs a fragment of the language of higher-order truth assessments which is translated into the meta-language of the assessment-sensitive relativist. This translation involves an extension of the expressive power of the English object language. Such an extension allows expressing the meta-linguistic claims in the object language. However, such an extension was never intended by the proponents of assessment-sensitive relativism. Moreover, this extension engenders a confusion between the object language and the metalanguage. The metalanguage of the assessment-sensitive relativist is a technical language that is not subject to the vagaries of natural languages. The translation between this metalanguage and the language of higher-order truth assessments—which is used to express the retraction data—treats the metalanguage as if it were part of our natural language, thus confusing the object language with the metalanguage.

Reply

The fact that such an extension was not intended by the proponents of assessment-sensitive relativism does not imply that it is illegitimate. The translation does not obliterate the difference between the role of a metalanguage of an empirical semantics and the role of its object language. What the translation accomplishes is just a way to express the higher-order assessments using the metalanguage of the assessment-sensitive relativist.

Objection 2

The language of higher-order truth assessments is not the language used by ordinary speakers: thus, interpreting the retraction data with this language is misinterpreting ordinary speakers.

Reply

I do not claim that ordinary speakers explicitly use the language of higher-order truth assessments. Rather, I claim that the truth conditions of sentences expressing retraction—e.g. 'I was wrong'—are equivalent to higher-order assessments. Assserting that there are such equivalences amounts to claiming that ordinary speakers performing retractions are rationally committed to these higher-order assessments *if* retraction data has the role that the assessment-sensitive relativist thinks it has.

To sum up, in this chapter I have formulated a challenge to MacFarlane's (2014) assessment-sensitive relativism: if we integrate assessment-sensitive relativism with the trumping framework developed

in Moruzzi and Wright (2009), Evans' challenge is effective against MacFarlane's account of assertion.

Of course, the assessment-sensitive relativist could happily reject the trumping framework. But until such a one gives a principled motivation for finding illegitimate the expressive resources of such a framework, the challenge remains open.

Notes

1. For a critique of possible-world truth relativism see Cappelen and Hawthorne (2009).
2. I have added an extra premise (step 2) to Greenough's formulation. See also McIntosh (2014) for an interesting reconstruction of Evans' dialectics.
3. See Kölbel (2002: 32, 91), Kölbel (2003: 69–72) and MacFarlane (2014: 103, 116).
4. I will focus here on assertion and I will not consider what a relativist should say about belief, though I find this latter issue a central question for relativism.
5. Here I focus on the assessment sensitivity thesis whereby the proposition expressed remains the same by changing the assessment context. If it is also a relative matter what content the assertion expresses, we have a further dimension of relativity. I will ignore this further dimension in the rest of the chapter.
6. The following norms are all intended to be about propositions as in MacFarlane's book. I'll use italics to signal that I am referring to propositions.
7. Fermat's Last Theorem is that no three positive integers a, b and c satisfy the equation $a^n + b^n = c^n$ for any integer value n greater than 2.
8. Teresa Marques has argued for a similar thesis in Marques (2014). My argument differs crucially from Marques' argument since I consider the normative setting of MacFarlane (2014), whereas Marques targets MacFarlane's previous normative setting presented in MacFarlane (2005).
9. Except for the assignment, but we are here assuming that the we are keeping fixed the language.
10. For another application of the trumping framework to the notion of moral progress, see Coliva and Moruzzi (2012).
11. Assuming that the truth of $\neg$*FERMAT* entails the falsity of *FERMAT*.
12. According to the reflexive rule, truth is a necessary but not sufficient condition for having permission for making the assertion.
13. The conjunction of (6) and (4).
14. (5) and (1).

References

Cappelen, H., and J. Hawthorne. 2009. *Relativism and Monadic Truth.* Oxford: Oxford University Press.

Coliva, A., and S. Moruzzi. 2012. Truth Relativists Can't Trump Moral Progress. *Analytic Philosophy* 53 (1): 48–57.

Evans, G. 1979. Does Tense Logic Rest on a Mistake? ms, printed in G. Evans (1985), *Collected Papers*, 343–363. Oxford: Oxford University Press.

Greenough, P. 2011. Truth-Relativism, Norm-Relativism, and Assertion. In *Assertion: New Philosophical Essays*, ed. J. Brown and H. Cappelen, 197–232. Oxford: Oxford University Press.

Knobe, J., and S. Yalcin. 2014. Epistemic Modals and Context: Experimental Data. *Semantics and Pragmatics* 7 (10): 1–21.

Kölbel, M. 2002. *Truth Without Objectivity.* London: Routledge.

Kölbel, M. 2003. Faultless Disagreement. *Proceedings of the Aristotelian Society* 54 (1): 53–73.

MacFarlane, J. 2005. Making Sense of Relative Truth. *Proceedings of the Aristotelian Society* 105: 321–339.

MacFarlane, J. 2014. *Assessment Sensitivity: Relative Truth and Its Applications.* Oxford: Oxford University Press.

Marques, T. 2014. Relative Correctness. *Philosophical Studies* 167 (2): 361–373.

McIntosh, J. 2014. Evans's Challenge to Temporalism. Technical report, UCL Working Papers in Linguistics.

Moruzzi, S., and C. Wright. 2009. Trumping Assessments and the Aristotelian Future. *Synthese* 166 (2): 309–331.

Knowing the Facts: A Contrastivist Account of the Referential Opacity of Knowledge Attributions*

Giorgio Volpe

> 'I know' is supposed to express a relation,
> not between me and the sense of a proposition (like 'I believe')
> but between me and a fact. (Wittgenstein, *On Certainty*, §90)

Ordinary speakers of English, but also philosophers acting in their professional capacity, often describe agents as 'knowing the facts of the matter', characterise the outcome of our epistemic dealings with the world as

*The problem addressed in this paper, arising as it does from a distinctively realist view of the world and our knowledge of it, would likely have struck the late Eva Picardi as an internal puzzle generated by some very questionable presuppositions. The reason why this has not deterred me from selecting it for discussion in this tribute to her memory lies in the fact that she would have considered the arguments offered in these pages with the open-mindedness, insight and generosity that have led so many of us to regard her as a model of intellectual integrity and a source of philosophical inspiration.

G. Volpe (✉)
University of Bologna, Bologna, Italy
e-mail: giorgio.volpe@unibo.it

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7_19

'knowledge of facts', and refer to what has successfully been established by inquiry as the 'known facts'. By using these and other such phrases, they seem to imply that knowledge is, at least in some cases, a relation to *facts*. However, in epistemological circles the type of knowledge that is canonically ascribed to agents by issuing statements of the form 'S knows that p' goes under the label of 'propositional knowledge'. And, although the qualifier 'propositional' is sometimes used rather innocently as just a way to register the form of such knowledge attributions, the view that propositional knowledge is a relation to *propositions* is currently taken for granted by many epistemologists. This may be due to the circumstance that the idea that knowledge (from now on I will omit the qualifier 'propositional') is a species of belief has largely survived the demise of the traditional analysis that equates it to justified true belief. For belief is standardly assumed to be a propositional attitude, i.e., a relation to propositions; and if knowledge is just an especially valuable form of belief, it too will have to be regarded as a relation to propositions. Moreover, the idea that knowledge is in many interesting respects similar to such mental states as belief, hope and fear is so entrenched, that even those philosophers that refrain from regarding it as an especially valuable form of belief tend to retain the idea that it is a propositional attitude (thus, for instance, Williamson 2000, 42 f.).

I myself am inclined to think that knowledge, unlike belief, is just what the above-mentioned phrases suggest, i.e., a relation to *facts*—where of course facts should not to be taken to be mere true propositions. In this paper, however, I will not attempt a full-fledged defence of this claim, but I will content myself with bringing some grist to the 'factualist' mill by, first, disentangling the claim that knowledge is a relation to facts rather than propositions from some controversial linguistic theses with which it is often associated, and then by sketching a fresh solution to a problem of referential opacity that needs to be addressed by anyone wishing to endorse it.

In Sect. 1 I briefly introduce the linguistic arguments with which Zeno Vendler and other philosophers have attempted to corroborate the view that I am concerned to defend. In Sect. 2 I present the main objection that has been raised against such arguments and disentangle the metaphysical claim that knowledge is a relation to facts rather than propositions from the linguistic claims with which such arguments end up burdening it. In Sect. 3 I present the problem of referential opacity that is raised by the metaphysical claim that knowledge is a relation to

facts rather than propositions and identify the minimal semantic thesis with which the metaphysical claim must be conjoined to generate the problem. In Sect. 4 I consider some possible solutions to the difficulty, including Vendler's own solution, a 'Moorean' solution and Keith Hossack's solution; my aim in this section is merely to review certain prima-facie reasons why it is worth looking elsewhere for a fully satisfactory resolution of the problem. So in Sect. 5 I put forth my own proposal, which harnesses the contrastivist idea that knowledge is a three-place relation between an agent, a fact, and a contrast term. In Sect. 6 I defend it against a pressing objection and take up the issue of the nature of the contrast term involved in the knowledge-relation. Finally, in Sect. 7 I review the main virtues of the solution and suggest that the circumstance that it falls out rather naturally from the core tenet of epistemological contrastivism may be regarded as providing indirect support for epistemological contrastivism itself.

1 Linguistic Arguments for Knowledge of Facts

The philosophers who maintain that knowledge is a relation to facts rather than propositions usually offer linguistic arguments in support of their view.[1] They argue that there is compelling linguistic evidence for the truth of three interrelated claims concerning the *that*-clauses employed in *oratio obliqua* and attitude reports. The first is that they are syntactic units. The second is that they behave as singular (i.e., referential) terms. And the third is that they are used to refer to different kinds of entities in different linguistic contexts—in particular, that in knowledge attributions of the form 'S knows that p' they are used to refer to *facts* rather than propositions. Each of these claims presupposes the previous one; but while the first is reasonably uncontroversial, the other two are considerably more contentious.

Together, the first two claims make up the so-called 'relational' or 'standard' account of *oratio obliqua* and attitude reports, which avers that the logical form of indirect speech reports like 'Bill said that electrons have a negative charge' and attitude reports like 'John believes that Husserl is an American philosopher' is genuinely relational, and that such reports are true just in case the referent of their subject stands in the relation denoted by their predicate to the referent of their embedded *that*-clause. The popularity of this account is mainly due to the fact

that it provides a simple and elegant account of the logical form of many manifestly valid inferences, such as

> John believes that Husserl is an American philosopher, and so does Greta.
> So, there is something that they both believe – to wit, that Husserl is an American philosopher.

> Harold believes everything that Bill says.
> Bill says that electrons have a negative charge.
> So, Harold believes that electrons have a negative charge.

According to the relational account, such inferences instantiate familiar inferential patterns of first-order logic, so their validity is easily explained. And several philosophers (including, e.g., Anderson 1984; Bealer 1993; Horwich 1998; Parsons 1993; Schiffer 2003; Künne 2003; Lynch 2009) maintain that this provides compelling reason not only for accepting the relational account, but for acknowledging the existence of *propositions*, in so far as these are conceived as the entities, *whatever they are*, to which reference is made by means of the relevant *that*-clauses.

The relational account is controversial, but the thesis that there is unambiguous linguistic evidence for the claim that *that*-clauses are used to refer to different kinds of entities in different linguistic contexts is even more so (Moffett 2003, 82–83; Betti 2015, chapter 5). Indeed, while the relational account may justifiably be regarded as the 'standard' account of such matters, the philosophers who take the linguistic data to warrant the further metaphysical claim that in some contexts *that*-clauses are used to refer to propositions and in others are used to refer to facts are just a small minority.

Back in the early seventies, it was Vendler (1972, chapter 5) who first maintained that that are two kinds of *that*-clauses, the 'subjective' and the 'objective'. He argued that *that*-clauses belonging to these two kinds have different co-occurrence restrictions—the subjective ones fitting subjective P-nouns and subjective verbs like 'say' and 'believe', the objective ones fitting objective P-nouns and objective verbs like 'tell' and 'know'—, as well as different transformational potential—the latter but not the former being open to *wh*-nominalization (replacement *salva grammaticalitate* by an appropriate indirect question clause). Vendler took *that*-clauses to be *noun*-clauses and concluded that, while subjective *that*-clauses refer to propositions, objective ones refer to facts,

contributing totally distinct kinds of 'objects' to belief and knowledge attributions. Around the same years, linguists Paul and Carol Kiparsky (1970) introduced a closely related distinction between 'non-factive' and 'factive' predicates and clauses, which has played a central role in subsequent attempts to offer linguistic evidence for the claim that knowledge is a relation to facts rather than propositions. Thus, for instance, Parsons (1993) has argued that the incongruity of the sentences that result from the attempt to quantify over the referents of *that*-clauses occurring in factive contexts with locutions appropriate to propositions is evidence that *that*-clauses are ambiguous, referring sometimes to propositions and sometimes to facts. Peterson (1997, 7) has maintained that '[s]emantic (and some syntactic) structural facts suggest that certain types of structures are typically used by speakers to refer to facts, and other types to refer to propositions'. And Holton (2017) has recently defended what he calls the 'Facts-for-factives conjecture' (partly) on the ground that it explains the absence of contra-factives in English and other Indo-European languages.

The contention of these philosophers is that there is strong linguistic evidence for the metaphysical conclusion that knowledge, unlike belief, is a relation to facts rather than propositions. Such evidence is alleged to come from the wide grammatical divergence between the verbs "know" and "believe", and between factive and non-factive linguistic contexts. But it should be borne in mind that the divergence, even if genuine, supports the metaphysical conclusion only on the assumption that *that*-clauses are indeed singular terms—an assumption which in recent years has come under the focus of heavy criticism.

2 Knowledge of Facts Without the Relational Account

Objections to the assumption that *that*-clauses occurring in *oratio obliqua* and attitude reports are singular terms usually turn on cases of substitution failure: if such clauses were indeed singular terms, they should always be substitutable *salva veritate* and *salva congruitate* with coreferential singular terms and natural language particular quantifiers restricted to propositions (and/or facts); but they cannot, so they are not singular terms (Bach 1997; McKinsey 1999; Moltmann 2003; Rosefeldt 2008; Betti 2015).

The cases of substitution failure that allegedly refute the claim that *that*-clauses are singular terms are not all equally compelling, and friends of the claim have argued that none is really fatal to their view (e.g., King 2001; Künne 2003). My impression is that their efforts are not completely successful, but for present purposes it is not necessary to take a definite stand on the issue. For even if the linguistic evidence that allegedly supports the metaphysical claim that knowledge is a relation to facts rather than propositions would be defeated if *that*-clauses were not singular terms, the deeper motivation behind this claim is arguably not linguistic. Vendler himself explicitly articulated the more fundamental metaphysical and epistemological considerations that constituted his basic motivation for endorsing it:

> What I believe or what I say may fit the facts, in which case it is true; or it may fail to fit the facts, in which case it is false. What I know, however, is the fact itself, not something that merely corresponds, or fails to correspond, to the facts. (Vendler 1972, 114)

That knowledge is a relation to facts rather than propositions is a claim that philosophers with realist leanings will naturally tend to endorse, especially if they are inclined to accept a fact-based version of the correspondence theory of truth. And it is a claim that can be endorsed independently of the theses that *that*-clauses are singular terms and that *that*-clauses occurring in knowledge attributions of the form 'S knows that p' are used to refer to facts. My suggestion is indeed that the fate of the claim had better be separated from that of these theses, even if dropping the relational account of *oratio obliqua* and attitude reports will require abandoning the project of corroborating it by the linguistic evidence cited by philosophers like Vendler, Peterson and Parsons.

It is worth noting, in passing, that separating the fate of the metaphysical claim that knowledge is a relation to facts rather than propositions from the relational account of *oratio obliqua* and attitude reports has the side benefit of pre-empting the Ramsey-inspired objection that, if *that*-clauses occurring in knowledge attributions of the form 'S knows that p' are singular terms used to refer to facts, they will suffer reference failure whenever the sentence embedded in them is false, making it difficult to explain how the knowledge attributions in which they are embedded can express a proposition (Williamson 2000, 43; see Ramsey 1927, 155). Discussing this point at greater length is, however, a task

for another occasion. Coming back to our main concern, in what follows I will not offer any further arguments for accepting the metaphysical claim that knowledge is a relation to facts rather than propositions, other than the observation that it coheres well with a realist metaphysics and a conception of truth as correspondence with (mind-independent) facts. Having disentangled this claim from the controversial linguistic theses with which it is often associated, what remains for me to do is just, as I said, to defend it from a potentially fatal objection by sketching a new solution to a problem of referential opacity that needs to be addressed by anyone wishing to endorse it. To this I now turn.

3 A Problem of Referential Opacity

At first sight, it might look as if no problem of referential opacity could arise once the metaphysical claim that knowledge is a relation to facts rather than propositions has been disentangled from the linguistic theses that have been discussed in the last two sections. And indeed, the metaphysical claim, taken in isolation, cannot yield any such problem. However, there is a considerably less controversial *semantic* thesis that upholders of the claim are apparently committed to accept, namely, the thesis that knowledge attributions of the form 'S knows that p' are true just in case the referent of 'S' stands in the knowledge-relation to the fact (if any) that makes the proposition expressed by 'p' true. This thesis is weaker than the linguistic theses discussed in the last two sections: knowledge attributions of the form 'S knows that p' may have these truth-conditions even if the *that*-clauses embedded in them are not used to refer to facts, and indeed even if they do not behave as singular terms. From a metaphysical point of view, the thesis that they are true just in case there is a specific fact to which the referent of their grammatical subject stands in the knowledge-relation may seem quite a commitment, but it is a commitment that realist-minded philosophers endorsing a fact-based version of the correspondence theory may be perfectly happy with. And from a semantic point of view, it seems rather minimal. But adding this minimal semantic thesis to the metaphysical claim that knowledge is a relation to facts rather than propositions immediately yields a problem of referential opacity. This can be easily seen by considering a well-worn example featuring Lois Lane and Superman/Clark Kent.

Suppose (i) that knowledge is a two-place relation between an agent and a fact, and (ii) that knowledge attributions of the form 'S knows that

p' are true just in case the referent of 'S' stands in the knowledge-relation to the fact (if any) that makes the proposition expressed by 'p' true. Now, it seems clear that the fact that makes the proposition expressed by 'Superman can fly' true is the fact that Superman can fly. And of course, the fact that makes the proposition expressed by 'Clark Kent can fly' true is the fact that Clark Kent can fly.[2] But if (iii) the fact that Superman can fly and the fact that Clark Kent can fly are one and the same fact, it immediately follows that, if 'Lois Lane knows that Superman can fly' is true, then 'Lois Lane knows that Clark Kent can fly' is also true—a conclusion that flies in the face of the fact that knowledge attributions of the form 'S knows that p' are manifestly referentially opaque.[3]

One might perhaps bite the bullet here by adopting a revisionary attitude towards our ordinary practices of knowledge attribution, but of course it would be much better to avoid having to regard as mistaken a large portion of these practices, and in what follows I will only attend to ways of addressing the problem that do not involve such a pervasive error-theory. But before we turn to the solutions suggested in the literature, and to the one I wish to propose, another comment is in order.

Since the thesis that knowledge attributions of the form 'S knows that p' are true just in case the referent of 'S' stands in the knowledge-relation to the fact (if any) that makes the proposition expressed by 'p' true plays a crucial role in generating the issue, one might be tempted to tackle it simply by dropping this thesis—or, rather, simply by dropping the right-to-left part of the biconditional, which is the really troubling bit. This of course would entitle one to declare without inconsistency that knowledge is a two-place relation between an agent and a fact *and* that knowledge attributions like 'Lois Lane knows that Superman can fly' and 'Lois Lane ignores that Clark Kent can fly' can be simultaneously true. But apparently it would also force one to maintain that ordinary knowledge attributions do not reflect in any obvious way the actual epistemic relations that obtain (or fail to obtain) between the agents and the facts they supposedly are about. As will become clear in due course, this unsavoury outcome does not mean that dropping the crucial conditional cannot be *part* of a satisfactory solution to the problem; but it does mean that it cannot be the whole solution, since it is reasonable to expect that a satisfactory solution will connect in a suitable way the thesis that knowledge is a relation to facts rather than propositions to our ordinary practices of knowledge attribution. The proposals I go on to consider all fit this expectation.

4 Some Solutions to the Referential Opacity Problem

An obvious solution to the problem raised by the referential opacity of ordinary knowledge attributions would be to endorse a 'Moorean', or 'propositional', view of the individuation of facts, that is to say, a view on which facts are as fine-grained as propositions (Moore 1953, 256; see also White 1970, chapter 4; Fine 1982, 56–57; Searle 1995, chapter 9). On such a view, the fact that Superman can fly and the fact that Clark Kent can fly will be two different facts. So there will be no reason to expect that, by standing in the knowledge-relation with the former, Lois Lane will thereby stand in the knowledge-relation with the latter, and hence no reason to deny that 'Lois Lane knows that Superman can fly' and 'Lois Lane ignores that Clark Kent can fly' can be simultaneously true. The drawback of this solution is that it almost inevitably leads to the conclusion that facts are not 'in the world'—that they are 'abstract', or 'ideal', entities (Betti 2015, §5.2). Elsewhere I have argued that the view that facts are abstract entities may be successfully married to a realist metaphysics and a correspondentist conception of truth (Volpe 2005, 339–345; 2012, §6.3). However, the solution suggested by the adoption of 'Moorean' criteria for the individuation of facts sits rather uncomfortably with the claim that facts are themselves the worldly entities to which true propositions correspond, or the worldly truth-makers of true propositions, and it is unlikely to appeal to those philosophers who are concerned to maintain that facts are the *relata* of the knowledge-relation because they construe the claim that we can have knowledge not just of true propositions, but of the world itself, as the claim that we can have knowledge of facts *in the world*.

A solution that is more likely to appeal to such philosophers is due to Keith Hossack (2007, 6–10). This solution, unlike the previous one, involves dropping the minimal semantic thesis which is jointly responsible for the problem of referential opacity that concerns us here; however, Hossack does have something interesting to say on the way our ordinary practices of knowledge attribution connect with the metaphysical side of the matter. For although he maintains that knowledge is ultimately a relation to (worldly) facts, he treats *that*-clauses occurring in knowledge attributions as singular terms used to refer to *contents*—that is to say, to the modes of presentation of the facts *of which* agents are said to have knowledge when they are said to know *that* things are thus and so. The *that*-clause in 'Lois Lane knows that Superman can fly' is thus

taken to refer to a specific mode of presentation of the fact *of which* Lois Lane is said to have knowledge when she is said to know *that* Superman can fly. In general, Hossack's view is that knowledge attributions of the form 'S knows that p' are true if and only if there is some mental act or state (a belief, a memory, a judgment or an experience) x whose content is that-p, and that-p is a mode of presentation of a fact f of which S has knowledge in virtue of x (Hossack 2007, 7). The key relation in this analysis is the relation that holds between an agent and a fact when the former *knows of* the latter: this is the only epistemic constituent of the facts reported by knowledge attributions of the form 'S knows that p'. So, while knowing *that p* entails knowing *of the fact that p*, knowing *of the fact that p* does not entail knowing *that p*: 'Lois Lane knows that Superman can Fly' and 'Lois Lane ignores that Clark Kent can fly' can be simultaneously true even if there is one and the same fact of which Lois Lane has knowledge, a fact which may be presented either as the fact that Superman can fly or as the fact that Clark Kent can fly, but of which she has knowledge only in virtue of a mental act or state that presents it under the former mode.

Ingenious as it is, Hossack's solution has been accused of lacking independent plausibility: it is arguably unsupported by inferential and linguistic data, and perhaps even in conflict with them (Textor 2011, 76–80). And of course the assumption that *that*-clauses are singular terms raises its own worries. While these problems are worthy of further discussion (Hossack 2011, 126–127), they seem serious enough to warrant considering alternatives.

Vendler's (1972, 115–116) own proposal turns on the claim that cognitively limited agents know facts, as it were, *in perspective*: two persons may of course know the same fact, but the one fact they both know will often appear to them 'in different perspectives' (ibid., 115). Thus, for instance, to rehearse one of his examples, the person who knows that Onassis married Jacqueline Kennedy and the person who knows that Onassis married the widow of the late President will know the same fact, but if neither of them knows that Jacqueline is the widow of the late President, that fact will appear to them in different perspectives—which is why the former can be correctly said to know that Onassis married Jacqueline Kennedy *and* ignore that he married the widow of the late President, while the latter can be correctly said to know that Onassis married the widow of the late President *and* ignore that he married Jacqueline Kennedy.[4] Drawing on his claim that *that*-clauses fitting

objective verbs like 'know' are open to *wh*-nominalization, Vendler further suggests that the one fact that the first person knows as the fact that Onassis married Jacqueline Kennedy and the second as the fact that Onassis married the widow of the late President can be referred to transparently by saying that what they both know is *whom Onassis married*.

Unfortunately, *wh*-nominalization will not help when the aim is to refer in a transparent way not to the fact that different agents know in different perspectives, but to the fact that one and the same agent knows in one perspective and ignores in another. For what is it that Lois Lane simultaneously knows and ignores when she knows that Superman can fly and ignores that Clark Kent can fly? Of course it is not *what Superman can do*, or *what Clark Kent can do*, but it is not *who can fly* either. So perhaps the *wh*-nominalization suggestion had better be dropped as unessential to the proposal. On the other hand, Vendler's solution is not as explicit as one might wish about the connection between the metaphysical claim that knowledge is a relation to facts rather than propositions and our ordinary practices of knowledge attribution. But what is needed to fill this gap is just an account of the way relations to perspectives are encoded in knowledge attributions, and a natural candidate for this role immediately comes to mind: one merely needs to construe Vendler's subjective 'perspectives' as linguistically expressible 'modes of presentation'. An appealing version of Vendler's solution will then involve the claim that knowledge is a three-place relation between an agent, a fact and a mode of presentation of the fact known by the agent—a mode of presentation that may be encoded in knowledge attributions of the form 'S knows that p' as the sense of (or: as the proposition expressed by) the sentence that follows the complementizer 'that'.

The claim that knowledge is a three-place relation between an agent, a fact and a mode of presentation is obviously incompatible with the thesis that knowledge attributions of the form 'S knows that p' are true just in case the referent of 'S' stands in the knowledge-relation to the fact that makes the proposition expressed by 'p' true, but the main virtue of this way of fleshing out Vendler's solution is perhaps that the semantics that it assigns to our ordinary knowledge attributions does not yield the unpalatable conclusion that 'Lois Lane knows that Superman can fly' and 'Lois Lane ignores that Clark Kent can fly' cannot be simultaneously true. On the other hand, talk of 'modes of presentation' of facts is unlikely to be to everyone's taste. And in the end, the picture turns

out to be so close to the one offered by those neo-Russellian accounts of propositional attitudes that invoke 'guises' or 'modes of presentation' of propositions to make one wonder whether it would be really the sort of picture that one could happily subscribe to. Again, these problems may be worthy of further discussion, but they seem serious enough to warrant considering alternatives. So let me finally turn to the solution that I believe may deliver the goods.

5 A Contrastivist Solution

The solution I have in mind shares with the fleshed-out version of Vendler's proposal the idea that knowledge is a three-place relation—not, however, a three-place relation between an agent, a fact and a mode of presentation, but a three-place relation between an agent, a fact and a *contrast*. That knowledge is a three-place relation involving a contrast is the distinctive claim of *epistemological contrastivism*, a view suggested by some early remarks of Fred Dretske (1970, 1972), which has then been championed by Jonathan Schaffer (2004, 2005), Walter Sinnott-Armstrong (2008) and Adam Morton (2012) among others.[5]

To get a sense of the view, consider two different ways in which a person could self-ascribe knowledge of the fact that Clyde sold his typewriter to Alex (Dretske 1981, 373). A person could claim to know that Clyde *sold* his typewriter to Alex, or that Clyde sold his typewriter to *Alex*. But it seems clear that a person who claims to know that Clyde *sold* his typewriter to Alex is not claiming to know the same thing as a person who claims to know that Clyde sold his typewriter to *Alex*. For of course, as Dretske noted, a person who knows that Clyde *sold* his typewriter to Alex must be able to rule out the possibility that he *gave* it to him, or that he *loaned* it to him, while a person who knows that Clyde sold his typewriter to *Alex* must be able to rule out the possibility that he sold it to *John* or to *Bill*. The moral suggested by the example is then that in knowledge attributions of the form 'I know that p', or, more generally, 'S knows that p', *intonational focus* can operate, in Schaffer's phrase, as a 'mechanism of contrastivity', i.e., as a mechanism that fixes the specific contrast relative to which an agent claims, or is said to possess, knowledge that p. There are other mechanisms that are apparently employed to similar effect in knowledge attributions of the form 'I know that p' or 'S knows that p', for instance *cleft construction*, as in 'Clark knows that it was Jane who gave a lift to Bill', or 'Clark knows that it was to Bill

that Jane gave a lift'. Again, it seems clear that a person who knows that it was Jane who gave a lift to Bill must be able to rule out the possibility that it was *Ann* or *Susan* who gave him a lift, while a person who knows that it was to Bill that Jane gave a lift must be able to rule out that she gave a lift to *Ronald* or to *Mike*. The thrust of epistemological contrastivism, however, is not merely that some knowledge attributions exploit mechanisms of contrastivity like intonational focus or cleft construction, but that knowledge attributions of the form 'S knows that p' *always* introduce a contrast, even when this is not encoded through a specific semantic mechanism, because it is the knowledge-relation itself that always involves a contrast term: 'To know that x is A is to know that x is A within a framework of relevant alternatives, B, C, and D' (Dretske 1970, 1022).

There are of course significant differences among the views of self-styled epistemological contrastivists. Epistemological contrastivism is sometimes advertised as the most compelling account of the data provided by ordinary language knowledge attributions (Schaffer 2004), other times as a view which is less concerned with the semantics of common epistemic claims than with distinctively epistemological issues (Sinnott-Armstrong 2008, 268). Most champions of epistemological contrastivism take it for granted that one of the terms of the knowledge-relation, namely, the 'thing' that is known, is a proposition, but at least one of them—besides, that is, the author of these pages—maintains that it is a fact (Morton 2012, 101–102; cf. Dretske 1970, 1022). And while some writers take the relevant contrast to be a proposition, usually a disjunctive one, others take it to be a class of propositions, or a class of alternative possibilities (Snedegar 2014).

I shall return to this last issue below; here I wish to note that there are apparently several good reasons for subscribing to the claim that (propositional) knowledge is a three-place relation with a contrast class (or proposition) among its terms. First, there are linguistic reasons. Many ordinary knowledge attributions are explicitly ternary, knowledge-*wh* attributions including an interrogative complement wear ternicity on their sleeve, and several linguistic tests arguably support the conclusion that even those knowledge attributions that look binary actually encode a relation to a question, which in turn is the sort of thing that typically presents a contrast (Schaffer 2005, 244–254). Second, there are 'genealogical' reasons. Knowledge attributions arguably serve the twofold goal of flagging reliable informants and scoring inquiry, and again both

functions are best served by encoding a relation to a question, which is the sort of thing that typically presents a contrast (Schaffer 2005, 236–239; Morton 2012). Third, there are epistemological reasons. The contrastive structure fits perception, which is essentially an ability to discriminate among alternative stimuluses or states of affairs (Schaffer 2005, 243). Moreover, epistemological contrastivism not only offers an account of the context sensitivity of binary knowledge attributions which is arguably more satisfactory than the account offered by its closest predecessor, epistemological contextualism, but uses it as the basis of an attractive treatment of Cartesian sceptical paradoxes, as well as of a plausible diagnosis of what goes wrong in Moore's argument for the existence of an external world (Schaffer 2004; 2005, 259–268). But the most compelling epistemological consideration in favour of contrastivism is perhaps, for those who believe that knowledge requires reasons, that reasons are themselves always 'reasons for one thing as opposed to another' (Sinnott-Armstrong 2008, 257),[6] which is why, in the end, it may seem inevitable to regard knowledge as a ternary relation.

This is obviously not the place to embark in a detailed discussion of the arguments that are advanced in support of epistemological contrastivism. For present purposes, however, there is apparently enough independent evidence going for it to justify careful consideration of its implications for the challenge that the referential opacity of ordinary knowledge attributions poses to the thesis that knowledge is a relation to facts rather than propositions.

Here then, at last, is the proposal. Knowledge is indeed a three-place relation between an agent, a fact (not a proposition!) and a contrast term, and the referential opacity of (apparently binary) knowledge attributions of the form 'S knows that p' arises from the circumstance that their utterances always involve an implicit contrast term, which can vary as a result of replacing the sentence following the complementizer 'that' with a different sentence—even if the propositions expressed by the two sentences correspond to (or: are made true by) the same fact.

This proposal, like Hossack's and the suggested revision of Vendler's own proposal, involves divorcing the metaphysical claim that knowledge is a relation to facts rather than propositions from the semantic thesis that knowledge attributions of the form 'S knows that p' are true just in case the referent of 'S' stands in the knowledge-relation to the fact (if any) that makes the proposition expressed by 'p' true. However, the divorce takes place in a way that does not require abandoning the

common wisdom that 'Lois Lane knows that Superman can fly' and 'Lois Lane ignores that Clark Kent can fly' can both be simultaneously true.

In Lois Lane's case, unlike in Dretske's typewriter case, focus plays no role: there is no one sentence to be read with the stress on different words. However, it is the very fact that what Lois Lane is taken to know is that *Superman* can fly, while what she is taken to ignore is that *Clark Kent* can fly, that points to the different contrasts that are involved in the relevant knowledge attributions (as well as in the corresponding epistemic states of the agent): for to be able to know that *Superman* can fly, one must be able to rule out the possibility that (say) *Superman* can only walk and run, or that *Superman* can only walk, run and swim, etc.; while to be able to know that *Clark Kent* can fly, one must be able to rule out the possibility that *Clark Kent* can only walk and run, or that *Clark Kent* can only walk, run and swim, etc. These are different contrasts, which is why Lois Lane can know that Superman can fly and at the same time ignore that Clark Kent can fly: since her capacity to rule out the former contrast need not be matched by a corresponding capacity to rule out the latter, nothing prevents her from standing in the knowledge-relation to the fact that Superman/Clark Kent can fly *relative to the former contrast* but not *relative to the latter*—which of course means that the knowledge attributions 'Lois Lane knows that Superman can fly' and 'Lois Lane ignores that Clark Kent can fly' can both be simultaneously true.

The point generalizes to all cases of referential opacity of knowledge attributions of the form 'S knows that p' that are potentially troubling for the claim that knowledge is a relation to facts rather than propositions. For such cases will invariably involve pairs of knowledge-attributions whose respective *that*-clauses differ from each other only for the substitution of one or more coreferential singular terms that the attributee would be unable to recognise as denoting one and the same thing.[7] And it is only to be expected that, whenever this condition is met, the agent's epistemic relation to the relevant fact will involve two different contrasts, only one of which he (or she) will typically be able to rule out.[8]

6 What Sort of Contrast?

There is a pressing question that needs to be addressed to make a plausible case that the contrastivist solution that has just been sketched can serve its purpose. The question concerns the contrast term of the knowledge-relation, the nature of which has been assumed to licence the

claim that there are indeed two distinct contrasts involved in Lois Lane's epistemic relation to the fact that Superman/Clark Kent can fly. Now, someone might object that, Superman and Clark Kent being one and the same person, if Lois Lane can rule out such alternatives as that Superman can only walk and run, then, by the same token, she will be able to rule out such alternatives as that *Clark Kent* can only walk and run; therefore, it will be impossible for her to stand in the knowledge-relation to the fact that Superman/Clark Kent can fly relative to the former contrast but not to the latter.

This objection can be easily answered—provided, that is, one is willing to embrace the plausible idea that what being able to rule out a possible alternative consists in is, basically, *having adequate evidence that it does not obtain*. For it seems clear that, ignoring that Superman and Clark Kent are one and the same person, Lois Lane may well have adequate evidence that the possibility, say, that *Superman* can only walk and run does not obtain, without thereby having adequate evidence that the possibility that *Clark Kent* can only walk and run does not obtain. The point could be put by saying that (epistemic) 'ruling out' contexts are referentially opaque. However, I do not want to suggest that this is so because which alternatives the evidence possessed by an agent counts against depends on the way the agent represents to himself (or herself) the alternatives in question, as if the ruling-out relation were a four-place relation between an agent, the evidence he (or she) possesses, a metaphysical possibility and a mode of presentation of that metaphysical possibility (if I did, one might reasonably wonder whether anything would be gained by preferring the contrastivist solution to the fleshed out version of Vendler's solution). Rather, what I wish to suggest is that the referential opacity of such contexts is evidence that the possibilities at stake are individuated more finely than both metaphysical possibilities and worldly facts, so that the possibility that *Superman* can only walk and run and the possibility that *Clark Kent* can only walk and run count as *distinct* possibilities.

I have described the issue in terms of alternative *possibilities*, and my point is in fact that the possibilities an agent must be able to rule out to count as knowing a given fact are really *epistemic* possibilities (in the sense of, e.g., DeRose 1991), i.e., possibilities that are individuated more finely than both metaphysical possibilities and worldly facts.

However, we saw in the previous section that not all epistemological contrastivists regard the contrast term of the knowledge-relation as a class of possibilities. Those that do not typically take it to be a proposition, usually a disjunctive one, or (equivalently: Snedegar 2014) a class of propositions. When the contrast term is taken to be a proposition or a class of propositions, that there are two distinct contrast terms involved in Lois Lane's epistemic relation to the fact that Superman/Clark Kent can fly is even more obvious, at least if propositions are individuated more finely than worldly facts (as they should be). So I conclude that, whether the contrast term is construed in terms of (epistemic) possibilities or of (classes of) propositions, the objection raised in this section can be successfully met.

7 Conclusive Remarks

I said at the beginning that the aim of this paper was just that of sketching a fresh solution to the problem that the referential opacity of knowledge attributions of the form 'S knows that p' poses for the claim that knowledge is a relation to facts rather than propositions. I have argued that this solution can be successfully defended against a pressing objection and will not pause to discuss criticisms aimed at epistemological contrastivism in general. Here I will just recap what I take to be the main virtues of the solution. The first virtue is of course that it does full justice to the circumstance that ordinary knowledge attributions like 'Lois Lane knows that Superman can fly' and 'Lois Lane ignores that Clark Kent can fly' can both be simultaneously true. Another virtue is that, unlike Hossack's and Vendler's proposals, the contrastivist solution does not invoke modes of presentation of facts, which makes it acceptable even to those who would rather avoid commitment to such entities. But the most interesting virtue of the solution is perhaps that it falls out rather naturally from an independently motivated thesis about the adicity of the knowledge-relation. This pre-empts the objection that postulating a contrast term of the knowledge-relation might be just a way to offer an ad hoc solution to the problem that has concerned us in these pages. On the other hand, showing that epistemological contrastivism comes with the extra bonus of a plausible solution to this problem may be a way to provide an indirect defence of its own core tenet.

Acknowledgements Previous versions of this paper were presented at the SeRiC Seminar of the Department of Philosophy and Communication Studies of the University of Bologna (21 June 2016), at the XII Conference of the Italian Society for Analytic Philosophy (Pistoia, 6 September 2016) and at the IX European Congress of Analytic Philosophy (Munich, 24 August 2017). Thanks to all present for discussion.

Notes

1. One notable exception is Hossack (2007), who seems to regard the claim that knowledge is a relation to facts (rather than propositions) as the default metaphysical option.
2. I am assuming that expressions of the form 'the fact that p', unlike (perhaps) *that*-clauses occurring in knowledge attributions, do name a fact. Even this is disputed (see, e.g., Rundle 1979; Betti 2015, 147–150), but upholders of fact-based versions of the correspondence theory of truth will obviously accept the assumption.
3. I ignore complications connected with molecular propositions and facts; if the reader thinks there are no molecular facts, I beg him (her) to construe everything I say as restricted to atomic propositions and facts.
4. This aspect of Vendler's view is ignored by Betti (2015, 177–179), who charges him with introducing a theory-laden term ('know') which is part and parcel of a technical language—a move she rightly criticises as being inconsistent with his own 'general strategy of taking language at face value' (ibid., 179), but which Vendler neither makes nor is committed to make.
5. The thesis that knowledge is a three-place relation involving a contrast may be taken to apply across the board, or it may be restricted to empirical knowledge ('knowledge of fact'). Here it is taken to apply only to empirical knowledge (a priori knowledge raises its own problems, which lie beyond the scope of this paper).
6. The claim that reasons are always 'reasons for one thing as opposed to another' may need some qualification. In any case, it seems to be exceptionlessly true when restricted to the ampliative (defeasible) reasons that are offered in support of the empirical claims that agents typically make about their world (recall that the sort of epistemological contrastivism discussed in this paper is restricted to empirical knowledge).
7. I ignore the possibility that substitution of logically equivalent sentences that the attributee would be unable to recognise as such might be troubling for the claim that knowledge is a relation to facts rather than propositions, for there is no reason to assume that, for any couple of logically equivalent sentences *s* and *t*, the fact that *s* just *is* the fact that *t* (and so no

reason to conclude that, for any couple of logically equivalent sentences *s* and *t* and any agent A, A stands in the knowledge-relation to the fact that *s* if and only if A stands in the knowledge-relation to the fact that *t*).

8. I say 'typically' because one can imagine cases in which the agent can rule out both contrasts on independent grounds. Such cases are no threat to the claim that knowledge is a relation to facts rather than propositions (they involve no failure of substitutivity).

References

Anderson, C.A. 1984. General Intensional Logic. In *Handbook of Philosophical Logic*, vol. II, ed. D. Gabbay and F. Guenthner, 355–386. Dordrecht: Reidel.

Bach, K. 1997. Do Belief Reports Report Beliefs? *Pacific Philosophical Quarterly* 78: 215–241.

Bealer, G. 1993. Universals. *The Journal of Philosophy* 90: 5–32.

Betti, A. 2015. *Against Facts*. Cambridge, MA and London: The MIT Press.

DeRose, K. 1991. Epistemic Possibilities. *The Philosophical Review* 100: 581–605.

Dretske, F. 1970. Epistemic Operators. *Journal of Philosophy* 67: 1007–1023.

Dretske, F. 1972. Contrastive Statements. *Philosophical Review* 81: 411–437.

Dretske, F. 1981. The Pragmatic Dimension of Knowledge. *Philosophical Studies* 40: 363–378.

Fine, K. 1982. First-Order Modal Theories III—Facts. *Synthese* 53: 43–122.

Holton, R. 2017. Facts, Factives, and Contra-factives. *Proceedings of the Aristotelian Society* 91: 245–266.

Horwich, P. 1998. *Truth*, 2nd ed. Oxford: Clarendon.

Hossack, K. 2007. *The Metaphysics of Knowledge*. Oxford: Oxford University Press.

Hossack, K. 2011. Replies to Comments. *Dialectica* 65: 125–151.

King, J. 2001. Designating Propositions. *The Philosophical Review* 111: 341–371.

Kiparsky, P., and C. Kiparsky. 1970. Fact. In *Progress in Linguistics*, ed. M. Bierwisch and K.E. Heidolph, 143–173. The Hague: Mouton.

Künne, W. 2003. *Conceptions of Truth*. Oxford: Clarendon.

Lynch, M.P. 2009. *Truth as One and Many*. Oxford: Oxford Unversity Press.

McKinsey, M. 1999. The Semantics of Belief Attributions. *Noûs* 33: 519–557.

Moffett, M.A. 2003. Knowing Facts and Believing Propositions: A Solution to the Problem of Doxastic Shift. *Philosophical Studies* 115: 81–97.

Moltmann, F. 2003. Propositional Attitudes Without Propositions. *Synthese* 135: 77–118.

Moore, G.E. 1953. *Some Main Problems of Philosophy*. London: Allen & Unwin.

Morton, A. 2012. Contrastive Knowledge. In *Contrastivism in Philosophy*, ed. M. Blaauw, 101–115. New York and London: Routledge.

Parsons, T. 1993. On Denoting Propositions and Facts. *Philosophical Perspectives* 7: 441–460.
Peterson, P.L. 1997. *Fact Proposition Event*. Dordrecht-Boston-London: Kluwer.
Ramsey, F.P. 1927. On Facts and Propositions. *Proceedings of the Aristotelian Society* 7: 153–170.
Rosefeldt, T. 2008. "That"-Clauses and Non-nominal Quantification. *Philosophical Studies* 137: 301–333.
Rundle, B. 1979. *Grammar in Philosophy*. Oxford: Clarendon.
Schaffer, J. 2004. From Contextualism to Contrastivism. *Philosophical Studies* 119: 73–103.
Schaffer, J. 2005. Contrastive Knowledge. In *Oxford Studies in Epistemology*, vol. 1, ed. T.S. Gendler and J. Hawthorne, 235–271. Oxford: Oxford University Press.
Schiffer, S. 2003. *The Things We Mean*. Oxford: Oxford University Press.
Searle, J.R. 1995. *The Construction of Social Reality*. New York: Free Press.
Sinnott-Armstrong, W. 2008. A Contrastivist Manifesto. *Social Epistemology* 22: 257–270.
Snedegar, J. 2014. Ethics and Contrastivism. The Internet Encyclopedia of Philosophy. http://www.iep.utm.edu/eth-cont/. Accessed 28 Sept 2016.
Textor, M. 2011. Knowing the Facts. *Dialectica* 65: 75–86.
Vendler, Z. 1972. *Res Cogitans*. Ithaca, NY: Cornell University Press.
Volpe, G. 2005. *Teorie della verità*. Milan: Guerini e Associati.
Volpe, G. 2012. *La verità*. Rome: Carocci.
White, A.R. 1970. *Truth*. London: Macmillan.
Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

Index

A
abstraction principles, 162, 163, 165, 168–170, 176–179
arithmetic, 74, 88–93, 98, 102, 108–110, 115–120, 125, 126, 128, 130–134, 145–149, 152–154, 163, 186, 195, 322
assertibility conditions, 169, 226, 229, 230, 236
assertion, 17, 19, 38, 39, 41, 75, 78, 79, 81, 135, 136, 154, 171, 220–223, 226, 229, 230, 234–236, 298, 358, 385–387, 389–397, 399
assessment-sensitivity, 362, 363, 388, 397
Austin, John L., 58, 82, 200, 233, 382

B
belief, 8, 9, 13, 18, 19, 23–25, 40, 44, 58, 154, 199, 230, 250, 265, 267–269, 272, 275–277, 297–310, 312, 313, 329, 332, 358, 367, 386, 399, 402, 405, 410
Bolzano, Bernard, 15, 19, 22, 31–35, 41–49, 51–63

C
Cappelen, Herman, 380, 381, 399
causes, 46, 248, 268, 276, 282–284, 286–292, 300, 302, 305
Chomsky, Noam, 3, 4, 320–327, 332, 333, 338, 339, 342, 344–350, 352–354
Cohen, Stewart, 359, 361, 380
content externalism, 241, 253, 255
context of assessment, 387–393, 397
context principle, 3, 68, 78, 122, 161, 163, 165, 173, 174, 177, 183–193, 195–200
contextualism, 4, 319, 361–370, 373–376, 379, 380, 387, 414
contrastivism, 403, 412–414, 417, 418

A. Coliva et al. (eds.), *Eva Picardi on Language, Analysis and History*,
https://doi.org/10.1007/978-3-319-95777-7

D
Davidson, Donald, 2–4, 265, 267, 270, 283, 284, 288–292, 320, 327–329, 338
definitions, 35, 36, 93, 95, 96, 98, 99, 102, 103, 105–111, 128, 129, 140, 142, 143, 146, 153, 155, 156, 162–165, 167, 171, 172, 175–178, 195, 199, 201, 349
deflationism, 379, 382
DeRose, Keith, 359–361, 365, 366, 379, 416
descriptions, 15–17, 22, 41, 68, 69, 71–73, 79, 80, 157, 170, 188, 228, 230, 233, 285, 289, 321
doxastic attitudes, 305, 313
Dretske, Fred, 255, 257, 412, 413, 415
Dummett, Michael, 2, 3, 7, 31, 57, 58, 68, 69, 78, 80, 82, 83, 164, 165, 170–179, 185, 187, 190, 196, 198, 200, 201, 216, 220, 226–229, 319, 322, 323, 331, 338, 354, 357

E
epistemic rationality, 266, 267, 270, 272, 276
Evans, Gareth, 256, 385–387, 389, 390, 394, 397, 399
explanation, 11, 92, 121, 166, 191, 193, 198, 271–273, 281–292, 341, 348, 355, 370, 372, 378
extensions of concepts, 92, 93, 96, 101, 126

F
facts, 11, 15, 69, 74–76, 82, 106, 123, 131, 169, 191, 192, 194, 200, 225, 226, 233, 235, 271, 284, 285, 325, 332, 391, 392, 401–411, 414–419
false, the, 57, 67, 68, 74–78, 81, 82, 96, 103, 127, 136–139, 143, 144, 149, 155, 156, 158, 198
Fantl, Jeremy, 342, 354, 380
Frege, Gottlob, 2, 4, 7, 17, 21, 22, 26, 31, 34–41, 46, 49, 51, 52, 54–59, 61, 68–83, 87–99, 101–111, 113–135, 137, 139–158, 161–163, 165, 167, 170, 173–175, 177–179, 183–196, 198–201, 206, 210, 214, 216, 220–223, 228, 231–233, 236, 241, 249, 256, 319, 357
  identity thesis, 32, 37–39, 42, 51, 54, 61

G
grounding, 19, 22, 32, 44, 46, 48, 49, 53, 59, 60, 62, 166, 178, 255

H
Hawthorne, John, 361, 380, 381, 399
higher-order truth assessment, 389, 390, 392–394, 398
Hossack, Keith, 403, 409, 410, 414, 417
Hume, David, 21, 93, 126, 151, 152, 162, 212, 313
Husserl, Edmund, 7, 9, 12, 14–22, 24–26, 31, 256, 403, 404

I
innatism, 348–351, 353
intentionality, 243, 245, 247, 255, 257
interpretation, 69, 99, 103, 153, 165, 171, 174, 175, 177, 184, 185, 187–190, 192, 193, 195–199, 201, 209, 214, 265, 282, 292, 330, 331, 354

introspective awareness, 242, 252, 253, 255, 257, 258
invariantism, 361–363, 378, 380
isomorphism, 23, 27

J
judgment, 17, 20, 21, 23–26, 71, 74, 75, 79, 81, 92, 120, 144, 175, 297, 301, 303, 313, 410

K
knowledge
  attributions, 342, 402, 403, 405–415, 417, 418
  of language, 319, 322–324, 333, 354

L
language, 2–4, 14, 20, 52, 53, 72–74, 80, 82, 92, 96, 123, 129, 150, 151, 164, 168, 169, 177, 183–187, 189, 191, 192, 194, 197, 199, 200, 205–207, 209, 210, 213, 215, 216, 221, 227, 292, 319–333, 337, 338, 341, 343–353, 355, 357, 374–376, 378, 388–390, 392, 393, 398, 399, 405, 413, 418
Lawlor, Krista, 271, 382
logicism, 87–89, 116–119, 125–128, 130–132, 134, 145–147, 151–153

M
MacFarlane, John, 362, 373, 380, 382, 386–391, 393, 395, 398, 399
manners of presentation, 242, 246, 247, 249, 256, 258
Mcgrath, Matthew, 380
meaning, 3, 4, 16–18, 20, 22, 35, 67–82, 103, 131, 161, 168, 174, 177, 183–193, 195, 198, 206, 207, 210, 211, 213–216, 219–221, 223, 224, 226–229, 232, 236, 284, 286, 288, 300, 319–321, 325–329, 331, 333, 349
Meinong, Alexius, 7, 8, 10–15, 22–27, 256, 313
metaontological minimalism, 166, 167, 170
metasemantics, 185, 191–194, 199–201
modes of presentation, 247–250, 256, 409, 411, 412, 417
Montminy, Martin, 381, 382

N
natural numbers, 88, 89, 93, 96–100, 102, 103, 106–110, 114, 127
norms, 197, 267, 282, 285, 306, 311, 333, 387, 399

P
pejoratives, 219–226, 228–233, 235
perspective, 106, 113, 120, 125, 149, 196, 200, 207–210, 215, 220, 225, 229, 230, 232, 234, 236, 302, 337, 354, 366, 393–397, 410, 411
phenomenal character, 241–243, 245, 251–258
phenomenological internalism, 241, 244
philosophy, 2–4, 7, 8, 11, 12, 14, 15, 21–23, 25–27, 31, 60, 62, 67, 162, 166, 171, 178, 179, 183, 184, 187, 191, 199, 205, 216, 241, 255, 256, 265, 270, 283, 284, 288–292, 297–299, 302, 320, 329, 337, 343, 346, 357, 418

Picardi, Eva, 2, 7, 25, 26, 31, 80, 148, 171, 174, 175, 177–179, 183–185, 188, 196, 201, 205, 206, 210, 216, 219–223, 225, 226, 229, 230, 232, 235, 236, 241, 256, 265, 270, 292, 297, 319, 324, 329, 332, 334, 357, 401
practical knowledge, 338–344, 352–354
propositional attitudes, 233, 245, 301, 412
propositional knowledge, 339, 342–344, 348, 352–354, 402
propositions, 12, 17–21, 26, 39, 41, 42, 45–47, 53–55, 61, 74, 81, 114, 118, 119, 127, 129, 131, 132, 134, 135, 147, 156, 158, 188, 220, 232, 233, 243, 298, 307, 339, 341, 342, 362, 385–387, 394, 399, 402–409, 411–415, 417–419
psychiatry, 282–284, 286, 288, 290–292

R
reasons, 11, 32, 39, 61, 110, 115, 142, 150, 158, 188, 206, 215, 242, 266, 268, 269, 271, 273, 282–292, 300, 310, 312, 323, 325, 330, 340, 342, 345, 353, 354, 403, 413, 414, 418
referential opacity, 402, 407, 409, 414–417
robust platonism, 164, 172
rules, 114, 123, 129, 140–142, 149, 155, 227–230, 338, 345–348, 350–353, 355
Russell, Bertrand, 2, 7, 8, 11–15, 17, 18, 21–26, 69, 71, 76, 77, 80–82, 89, 91, 99, 113, 114, 177, 183

S
Schaffer, Jonathan, 370, 381, 412–414
semantic individualism, 319, 320, 326, 332
semantic normativity, 332, 333
semantics, 68, 71, 78, 79, 82, 124, 127, 129, 130, 148, 151, 157, 163, 166, 169, 174, 177, 184, 194, 200, 206, 207, 211, 363, 378, 387–392, 398, 411, 413
sense and tone, 3, 220
Stanley, Jason, 339, 342–344, 353, 354, 361, 370, 380, 381
success, 176, 266, 276, 277, 308–310, 366
Szabo, Zoltan, 370, 381

T
thin objects, 166–170, 178
thin reference, 170, 173, 178, 179
tolerant reductionism, 170, 172, 176, 177, 179
tone, 12, 205–211, 213–216, 220, 221, 226, 228, 232, 236, 302
true, the, 67, 68, 74–76, 78, 81–83, 95, 96, 103, 127, 136–139, 143, 144, 149, 155, 156, 158
  true-because-of principle, 32, 44–46, 58
  true-iff principle, 32, 45, 58
trumping, 390, 392–399
truth
  ascriptions, 34, 41, 42, 45–47, 60
  relativism, 386, 387, 394, 399

V
Vendler, Zeno, 402, 404, 406, 410–412, 416–418

W
Weatherson, Brian, 380
Williamson, Timothy, 220, 226–228, 339, 342, 344, 354, 358, 380, 382, 406
Wittgenstein, Ludwig, 2, 12, 18, 23, 75, 76, 80–83, 113–115, 119, 121, 147, 150, 155, 183, 187, 188, 333, 337–339, 351, 352, 355