The Western Ontario Series in Philosophy of Science 80

Sorin Costreie

Early Analytic Philosophy – New Perspectives on the Tradition

Springer

Early Analytic Philosophy – New Perspectives on the Tradition

THE WESTERN ONTARIO SERIES IN PHILOSOPHY OF SCIENCE

A SERIES OF BOOKS IN PHILOSOPHY OF MATHEMATICS AND NATURAL SCIENCE, HISTORY OF SCIENCE, HISTORY OF PHILOSOPHY OF SCIENCE, EPISTEMOLOGY, PHILOSOPHY OF COGNITIVE SCIENCE, GAME AND DECISION THEORY

VOLUME 80

More information about this series at http://www.springer.com/series/6686

Sorin Costreie
Editor

Early Analytic Philosophy – New Perspectives on the Tradition

Springer

Editor
Sorin Costreie
Faculty of Philosophy,
Department of Theoretical Philosophy
University of Bucharest
Bucharest
Romania

ISSN 1566-659X ISSN 2215-1974 (electronic)
The Western Ontario Series in Philosophy of Science
ISBN 978-3-319-24212-5 ISBN 978-3-319-24214-9 (eBook)
DOI 10.1007/978-3-319-24214-9

Library of Congress Control Number: 2015950011

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

In memory of Jaakko Hintikka

Preface and Acknowledgments

This collection originates from a series of three conferences under the auspices of the *Bucharest Colloquium in Analytic Philosophy*. The first took place in 2010, and it was called *The Actuality of the Early Analytic Philosophy*. It was followed in 2011 by *Frege's Philosophy of Mathematics and Language* and in 2012 by *Philosophy of Mathematics Today*.

Most of the papers in the volume were presented, in an earlier version, at one of these conferences; the remaining essays were specially commissioned. In putting together the present volume, I have kept the focus on the original idea of the first conference, namely the contemporary relevance of early analytic philosophy. Although the papers of this collection are mainly dedicated to a specialized public, familiar with the issues and methods of analytic philosophy, the volume is designed so that it can also serve as a useful companion for various introductory courses covering the origin and the evolution of the analytic tradition.

All *Bucharest Colloquiums* took place at the University of Bucharest and were supported, at the institutional level, by the university's Department of Theoretical Philosophy and its Center for Logic, History and Philosophy of Science, and by the Romanian Society for Analytic Philosophy. I am indebted to the Faculty of Philosophy of the University of Bucharest for making it possible for me to co-organize this series of colloquiums. At a more personal level, I am grateful to all of the *Bucharest Colloquium's* participants, to my co-organizers—Mircea Dumitru and Gabriel Sandu—and to the many students who helped ensure that the conferences ran smoothly.

I would like to thank first and foremost the authors for their willingness to contribute to this collection and for their understanding and forbearance in seeing this project through. Matthias Schirn deserves special thanks, for in addition to his essay, he contributed valuable ideas to the project and attracted several valuable contributors. I also thank Lucy Fleet, my editor at Springer, and William Demopoulos, the managing editor of *The Western Ontario Series in Philosophy of Science*, for their constant assistance and support. I am grateful to Bill for all the philosophical guidance he gave to me during my graduate years at The

University of Western Ontario. My interest in analytic philosophy began during my undergraduate studies at the University of Bucharest with the Frege course of Sorin Vieru; it was consolidated by Adrian Miroiu's course on philosophical logic; and it was established as a constant direction in my academic life by Bill's graduate courses on Frege, Russell, Carnap and Co. The appearance of this volume in a series edited by Bill was a happy coincidence, for Lucy's suggestion that it be placed in the *Series* was given without prior knowledge of my acquaintance with its managing editor: The present has its roots deep in the past…

Bucharest Sorin Costreie
July 2015

Contents

Contributors

Sorin Bangu University of Bergen, Bergen, Norway

Patricia Blanchette University of Notre Dame, Notre Dame, USA

Derek H. Brown Brandon University, Brandon, Canada

Otávio Bueno University of Miami, Coral Gables, USA

David Fisher Indiana University, Bloomington, USA

Sébastien Gandon Blaise Pascal University, Clermont-Ferrand, France

Brice Halimi Université Paris Ouest Nanterre La Défense (IREPH) & SPHERE, Paris, France

Jaakko Hintikka Boston University, Boston, USA

Gregory Landini University of Iowa, Iowa City, USA

Gregory Lavers Concordia University, Montreal, Canada

James Levine Trinity College Dublin, Dublin, Ireland

Bernard Linsky Department of Philosophy, University of Alberta, Edmonton, Canada

Danielle Macbeth Haverford College, Haverford, USA

Charles McCarty Indiana University, Bloomington, USA

Thomas Meier Ludwig Maximilians University Munich, Munich, Germany

Marco Panza CNRS, IHPST, University of Paris 1, Panthéon-Sorbonne, Paris, France

Graham Priest University of Melbourne, Melbourne, Australia; CUNY Graduate Center, New York, USA

Matthias Schirn Ludwig Maximilians University Munich, Munich, Germany

Gheorghe Ştefanov University of Bucharest, Bucharest, Romania

Iulian D. Toader University of Bucharest, Bucharest, Romania

Introduction

Early Analytic Philosophy (spanning the period from 1879 to the early 1930s) is now known and regarded as a distinct philosophical tradition. The growing interest in this not fully explored domain of study is not just historical. In the last decades, new connections, interpretations, and ideas discovered in relation to it have shed a different light on the origins of analytic philosophy. What we have inherited from the analytic tradition is similar to the inheritance Frege told his son Alfred was contained in the unpublished works he would be leaving him: "Even if they are not pure gold, there is gold in them."

The present volume is a collection of papers that focus on discussing (g)old ideas mainly originating in the works of some of the central figures who initiated the analytic tradition: Gottlob Frege, Bertrand Russell, Ludwig Wittgenstein, and Rudolf Carnap. The central point of the book is to show how these ideas remain present and influential in current philosophical debates. The collection is the product of the need to better understand these ideas by placing them in their original setting and to systematically examine how these ideas might illuminate debates that animate current philosophical discussion.

The authors approach some crucial ideas of the founders of analytic philosophy with a keen interest in showing how much contemporary philosophy is indebted to its original setting; they also examine the extent to which current debates echo the "original" ones. The collection is designed to be a useful tool for those who recognize the fruitfulness of (g)old thoughts and their significant influence in current philosophical disputes.

The collection contains 19 new and original essays, written by junior and senior scholars in analytic philosophy who have been invited to write on what they take to be "old ideas in new clothes," ideas which still shape current philosophical debates. As one might expect, a rich and diverse picture emerges. The present collection covers many kinds of topics belonging to early analytic philosophy, topics which continue to intrigue analytic philosophers today.

The volume is organized into five parts. Each of the first four parts is dedicated to one of Frege, Russell, Wittgenstein, or Carnap. The last part gathers together several essays which discuss either the relation between two or more analytic

thinkers, or various important concepts such as the principles of abstraction and non-contradiction.

Some information about each of the contributions completes this introduction.

Patricia Blanchette explores the difficulties posed for Frege's claim that mathematical theories are collections of *thoughts* and that scientific continuity turns on thought-identity, by the conceptual development canonically involved in mathematical progress. Blanchette argues that the difficulties apparently posed to Frege's central views stem from an overly simple view of Frege's understanding of mathematical objects and of reference. The positive view recommended is one on which Frege's view of mathematical theories is largely consistent with, and helps make sense of, the phenomenon of theoretical unity across conceptual development.

Jaakko Hintikka draws attention to an important aspect of quantifiers, which was overlooked by Frege and others Fregean commentators like Kripke, namely their role of expressing, by their formal dependence on each other, the actual dependences between variables bound to them. The resulting flaw in Frege's and other logicians' logic began to be corrected only in IF logic. In any adequate logic, a fixed mode of identification is presupposed. The frameworks of identification can be perspectival or public. Hintikka claims that Kripke makes the same mistake about quantifiers as Frege and in addition assumes that only perspectival identification is needed in the last analysis. He also overlooks dependence relations between modal operators and quantifiers.

Danielle Macbeth discusses the model of a science that is outlined in Aristotle's *Posterior Analytics*. Frege is seen as one of the last great defenders of the model and a key figure in the very developments that have been taken to spell its demise. However, Macbeth claims that even though Frege remains true to the spirit of the model, he also modifies it in very fundamental ways. So modified, Macbeth suggests, the model continues to provide a viable and compelling image of scientific rationality by showing, in broad outline, how we achieve, and maintain, cognitive control in our mathematical investigations.

Matthias Schirn analyzes both the status and the role of Hume's Principle in Frege's logicist project. Schirn carefully considers the options that Frege might have had to establish the analyticity of Hume's Principle, bearing in mind that with its analytic or non-analytic status, the intended logical foundation of cardinal arithmetic stands or falls. Schirn reconstructs in modern notation essential parts of the formal proof of Hume's Principle in Frege's *Grundgesetze*. Schirn also scrutinizes Frege's characterization of abstraction in *Grundlagen*, §64, and criticizes in this context the currently widespread use of the terms "recarving" and "reconceptualization" by Crispin Wright and other neo-logicists. Schirn concludes his essay with some interesting reflections on the introduction of the cardinals and the reals by abstraction in the context of Frege's logicism.

Derek Brown is interested in what remains of the early analytic approach to perception—sense-datum theory—when it is both (a) divorced from an overly ambitious commitment to the idea that perception delivers a wealth of certain knowledge about what is perceived and (b) updated to accommodate phenomena in contemporary perceptual theory such as the *perceptual constancies*. Brown argues

that to achieve (a) one should "deflate" Russell's notion of *acquaintance* and (b) one can utilize the space created by deflated acquaintance knowledge to allow perceptual *representation* to resolve the ambiguities inherent in perceptual constancies. Brown thus offers a two-factor (acquaintance-representation) sense-datum theory to meet these challenges.

Gregory Landini points out some of the most striking intellectual differences between Whitehead and Russell that are relevant to *Principia*. Thus, Landini takes up the issues of typical ambiguity, the nature of classes, geometry, and the existence of mind and matter. It may seem surprising that there are such striking differences between Whitehead and Russell given their philosophical collaboration and very close personal relationship. Russell's neutral monism took an eliminativistic stance with respect to *life*, *mind, matter*, *motion*, *time*, and *change*. On the other hand, Whitehead viewed nature as an organically integrated *whole* within which *life*, *mind*, and *matter* have a genuine reality.

James Levine distinguishes three periods in Russell's philosophical development: the Moorean period, following his break with Idealism around 1899 through his attending the Paris conference in August 1900 at which he saw Peano; the period following the Paris conference through his prison stay in 1918; and his post-prison period, in which he becomes concerned with the nature of language as such. Levine argues that while the topic of vagueness becomes an explicit theme in his post-1918 writings, his view that ordinary language is vague plays a central role in his post-Peano practice and characterization of analysis. Levine claims that the failure to recognize the character of Russell's post-Peano conception of analysis reflects a broader misunderstanding of the character of Russell's philosophy, and of his place in the history of analytic philosophy.

Bernard Linsky understands the early chapters of *Principia Mathematica* as the result of a slow and laborious development out of Peano's original ideas. Linsky illustrates this development by studying a theorem that is not proved in those early chapters of *Principia*: Peirce's Law—$[(p \supset q) \supset p] \supset p$. Linsky distinguishes three Russellian systems of propositional logic: the first in *Principles of Mathematics* (1903), then the second in "The Theory of Implication" (1906), and the third in *Principia Mathematica* (1910). Linsky's paper is designed as an investigation of the role of Peirce's Law through those systems. Linsky shows that this valid formula is not even a theorem in the 1910 system although it is one of the axioms in 1903 and is proved as a theorem in 1906. The paper helps also to reconstruct some of the history of axiomatic systems of classical propositional logic.

Sorin Bangu focuses on the lectures on the philosophy of mathematics delivered by Wittgenstein in Cambridge in 1939. He discusses several lectures, the emphasis falling on understanding Wittgenstein's views on the most important element of the logicist legacy of Frege and Russell, the definition of number in terms of classes—and, more specifically, by employing the notion of one-to-one correspondence. Since it is clear that Wittgenstein was not satisfied with this definition (and with the overall logically oriented approach), the aim of Bangu's essay is to propose a reading of the lectures able to clarify why that was the case. This reading draws

connections between Wittgenstein's views on language and mind (expressed mainly in *Philosophical Investigations*) and his conception of mathematics.

Sébastien Gandon compares Wittgenstein's discussion of color exclusion in his "Some Remarks on Logical Form" to William E. Johnson's doctrine of determinable and determinate expounded in his *Logic*. Gandon's point in doing this comparison is not to uncover a hidden influence of Johnson on Wittgenstein. Gandon holds that instead of considering "Some Remarks on Logical Form" as a step in the journey from the *Tractatus* to the *Investigations*, we should see it as an integral part of a discussion, witnessed by Johnson (1921), which took place in Cambridge in the twenties. Gandon's paper ends by putting Wittgenstein's ideas in a broader historical context.

Brice Halimi focuses on Wittgenstein's characterization of logical truth in the *Tractatus*. Wittgenstein describes the general validity of logical truths as being "essential," as opposed to merely "accidental" general truths. Few commentators have focused on this point, and most of them have construed it as the claim that generalized propositions cannot be but contingently true, if true. Halimi's aim is to elucidate the crucial concept of essential general validity (a concept which brings into play the whole Tractarian conception of logic) and to explain that it has to do with a certain kind of generality, before any kind of necessity.

Charles McCarty and David Fisher provide a mathematical demonstration that the assumption that the logical theory of *Tractatus* yields a foundation for (or is at least consistent with) the conventional logics represented in standard propositional, first-order predicate, and perhaps higher order formal systems is false according to a preferred account of argument validity. McCarty and Fisher show that the hierarchy of variables—and, hence, of propositions—defined at 5.501 has the expressive power of (at least) finitary classical propositional logic. Also, they prove that Wittgenstein's hierarchy of iterated N-propositions, as specified in Remark 6, does not collapse: At any level k, one finds propositions at $k + 1$ or above that are not logically equivalent to any proposition formed at k or below.

Gheorghe Ştefanov focuses on Wittgenstein's "Private Language Argument," holding that no experience, conceived as an inner episode to which only the subject having it has direct access, can be semantically relevant. Stefanov sees that a direct consequence of this point is that no experience, thus conceived, can be epistemically relevant, and thus then, the traditional empiricist project is in danger. However, on one hand, McDowell seems to offer a solution, and, on the other hand, Stefanov claims that a different solution to the same difficulties might have been suggested by Wittgenstein himself in *On Certainty*.

Otávio Bueno discusses Carnap's logicist stance and critically evaluates three moves made by Carnap to accommodate the abstract mathematical objects within his empiricist program: (i) the "weak logicism" in the *Aufbau*; (ii) the combination of formalism and logicism in the *Logische Syntax*; and (iii) the distinction between internal and external question characteristic of Carnap's involvement with modality. The outcome of Bueno's paper is an interesting picture of the clear interplay between Carnap's philosophy of science and his work in the philosophy of mathematics, and the developments of his ideas along these lines.

Greg Lavers examines the fundamental views on the nature of logical and mathematical truth of both Frege and Carnap. Lavers argues that their positions are much closer than is standardly assumed. Lavers also argues against the common point that Frege was interested in analyzing our ordinary mathematical notions, while Carnap was interested in the construction of arbitrary systems, for, as Lavers claims, our ordinary notions play, in a sense, an even more important role in Carnap's philosophy of mathematics than they do in Frege's. Lavers' paper ends by rejecting Tyler Burge's interpretation of Frege which is in opposition to any reasonable Carnapian reading of Frege.

Thomas Meier aims to provide a historical reconstruction of the interconnections between Carnap's *Aufbau*, Kuhn's model of theory change, and the structuralist view of scientific theories. Meier claims that scientific structuralism is rooted in Carnap's early work, especially in the *Aufbau*. Meier claims that Carnap's idea of purely structural definite descriptions exposed in the *Aufbau* can be seen to be analogous with the goal of the structuralist view of representing our knowledge about scientific theories structurally. He also discusses how the development of the structuralist view is strongly motivated by Kuhn's conception of theory change.

Marco Panza is concerned with the relation between analyticity and what he calls "epistemic economy." This relation is analyzed in the context of questioning the analyticity of abstraction principles. Panza claims that one important virtue of these principles that is commonly overlooked is their epistemic economy. Thus, most of the paper is dedicated to a careful and interesting examination of this notion in the context of defining real numbers.

Graham Priest provides an analysis and commentary on the second half of Jan Łukasiewicz's book *On the Principle of Contradiction in Aristotle* (1910). The book contained a critique of the traditional attitude to the principle of non-contradiction and a re-evaluation of its significance in light of contemporary developments in logic. Priest shows that Łukasiewicz is seen to be badly torn, for even though he eventually endorses the principle, he does so, not in virtue of the evidence he considers, but despite it. In particular, he considers arguments against the principle, drawn from Hegel, Meinong, and the paradoxes of self-reference.

Iulian Toader discusses Weyl's criticism of Dedekind's principle that there is no scientific provability without proof. This criticism, Toader claims, challenges not only a logicist norm of belief in mathematics, but also a realistic view about whether there is a fact of the matter as to which norms of belief are correct.

Sorin Costreie

Part I
Frege

Frege on Mathematical Progress

Patricia Blanchette

Introduction

One of the central motivations behind Frege's concern with *thoughts* is his concern with the communal nature of science. The fact that people separated by gulfs of time, of space, and of language can share a common science is, from Frege's point of view, due to the fact that the substance of a given science is not a collection of sentences or of ideas, but of thoughts, the kinds of things that can be expressed by sentences of different languages, and can be conveyed from person to person despite differences in ideas or contingent circumstance. As Frege himself puts it,

> Can the same thought be expressed in different languages? Without a doubt, so far as the logical kernel is concerned; for otherwise it would not be possible for human beings to share a common intellectual life.[1]

In addition to the *communal* nature of science, Frege is also importantly concerned with its developmental side, i.e. with the fact that sciences, mathematical ones in particular, experience significant conceptual refinement over time. This

Versions of this essay were presented at the 2014 "Frege@Stirling" workshop at Stirling University, and at the 2014 Logic Colloquium in Vienna. Many thanks to the organizers and audience members, especially to Philip Ebert, Bob Hale, and Rob Trueman for helpful comments.

[1]Frege (1979) 6, (1983) 6 from the "Logic" notes, undated. See also Frege (1892a) 29 (160/146), and Frege (1892b) 196 note (185/170) for similar sentiments about the sharing of thoughts as the ground of common science.

P. Blanchette (✉)
University of Notre Dame, Notre Dame, USA
e-mail: patricia.blanchette.1@nd.edu; blanchette.1@nd.edu

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_1

circumstance is especially important from the Fregean point of view in mathematics, since he takes his own project, one that involves highly non-trivial reconceptualizations of central mathematical notions, to be of a piece with the history of conceptual development in mathematics generally. On the importance of conceptual development in mathematics, Frege says at the beginning of *Grundlagen*:

> After deserting for a time the old Euclidean standards of rigor, mathematics is now returning to them, and even making efforts to go beyond them. … The discovery of higher analysis only served to confirm this tendency; for considerable, almost insuperable, difficulties stood in the way of any rigorous treatment of these subjects … The concepts of function, of continuity, of limit and of infinity have been shown to stand in need of sharper definition. Negative and irrational numbers, which had long since been admitted into science, have had to submit to a closer scrutiny of their credentials. [Frege (1884) §1]

And as to the connection between his own work and this tradition:

> In all directions these same ideals can be seen at work – rigour of proof, precise delimitation of the concept of validity, and as a means to this, sharp definition of concepts. (*...die Begriffe scharf zu fassen*.)
>
> Proceeding along these lines, we are bound eventually to come to the concept of Number, and to the simplest propositions holding of positive whole numbers… [Frege (1884) §§1–2]

The last-mentioned project, that of providing a deeper analysis of the concept of Number, and of "the simplest propositions holding of positive whole numbers," is the central work of *Grundlagen*.

One of the crucial features of conceptual analysis in mathematics, as Frege sees it, is that it is often highly non-trivial:

> Often it is only after immense intellectual effort, which may have continued over centuries, that humanity at last succeeds in achieving knowledge of a concept in its pure form, in stripping off the irrelevant accretions which veil it from the eyes of the mind. [Frege (1884) p. vi]

But now we seem to face a real difficulty. Fregean *thoughts* are not obviously the kinds of things that can survive the sort of significant conceptual development of which the history of mathematics consists. And if they cannot do so, then Frege's fundamental way of understanding the nature and the continuity of mathematical sciences is in tension with his conception of mathematical progress. Our central question in what follows is that of how we are to understand this tension in Frege, and of whether there is a plausible Fregean account of the nature of mathematics that makes sense both of continuity and of significant conceptual change over time.

Sense Versus Conventional Significance

One way of trying to clarify Frege's conception of the *sense* of an expression is by means of what a speaker of the language is aware of when, and in virtue of which, he or she is competent with respect to that expression. If this is the correct way to

understand sense, then the tension between Frege's view of continuity and his view of mathematical progress is stark: there is no sense in which, for example, a speaker's linguistic competence in the mid-18th century with the term "continuous function" requires any inkling of the content of its 19th-century analysans.

But as Tyler Burge has argued, the identification of Frege's notion of sense with linguistic meaning is a mistake.[2] Because the sense of a sentence is the fundamental truth-bearer, it is determined by the world in ways that can often outstrip the thin collection of information awareness of which is required for linguistic competence. As a corollary, the sense of an individual term can often be considerably richer than the collection of information that would be conveyed by a good dictionary. And as Burge has further argued, the separation of the "conventional significance" of a word—i.e. the material whose grasp constitutes linguistic competence—from the sense of that word offers a straightforward route to understanding how sense can outstrip what even expert speakers associate with a term or sentence.[3] In cases in which the conventional significance of a word is insufficiently precise to pin down a particular reference, the sense of that word, as determiner of reference, must go beyond that ordinary significance. As Burge sees it, the "extra" input is delivered, in the case of mathematics, by the mathematical facts themselves, those facts in whose systematization and explanation the term plays a central role. Concerning the example of the term "Number" as used prior to Frege's work, and hence whose conventional significance involves in Frege's view a good deal of imprecision, Burge asks:

> How could the term 'Number' indicate a definite "concept" when all current mathematical understanding and usage failed to determine a sense or concept? [Burge (1984) 10]

And replies:

> To say, as Frege says, that 'Number' *does* denote a concept and *does* express a sense is to say that the ultimate foundation and *justification* of mathematical practice supplements current usage and understanding of the term in such a way as to attach it to a concept and a sense. From this point of view, vague usage and understanding do not entail vague sense-expression. [Burge (1984) 11]

As we might put it, the rich body of mathematical facts that underpins our mathematical practice serves, together with the sometimes-incomplete information conventionally associated with mathematical terms, to fix determinate sense and reference on those terms. On this way of understanding the project, it is clear why the conceptual analysis so essential to mathematical progress requires non-trivial mathematical work: to gain clarity about the nature of the objects and concepts we have all along been referring to is, in part, to gain clarity about their mathematical properties and relations to one another. In gaining this clarity, we make it clear what it is in virtue of which it is *this* rather than *that* function or object to which we have been referring all along.

[2]See Burge (1979).

[3]See Burge (1984).

But while this clarification of the connection between conventional significance, mathematical facts, and sense provides an answer to Burge's question of how vagueness of conventional significance can be compatible with expression of determinate sense, it will not solve the whole of the difficulty sketched above, the difficulty of reconciling Frege's view of theoretical continuity as requiring thought-identity with the non-trivial nature of conceptual development. For it is not always the case that the mathematical facts underlying a given mathematical practice are sufficiently rich to disambiguate its terms in the context of later development. Suppose we have two mathematicians separated by a time-span in which there has been significant conceptual change, so that the later mathematician uses a mathematical term *t* whose sense and reference are fixed by a precise definition given in the interval. Burge's account gives us a way of understanding what has happened if the earlier mathematician uses the term *t* in such a way that her practice and the underlying mathematical facts together fix just that sense and reference, so that the conceptual work in the interval has consisted in clarifying a sense that was already determinate. But conceptual progress in mathematics is not always so simple. Consider for example the case of *continuous function*, a notion familiar to mathematicians before the precision instituted in the 19th century, and a notion open then to multiple non-equivalent precisifications. A plausible reading of the history is one on which nothing about eighteenth- or early-nineteenth-century practice served to pick out precisely one of the later-disentangled notions of *epsilon-delta continuity, uniform continuity,* and *differentiability.* Similarly for *cardinality*: pre-Cantorian practice with finite sets fails to pin down a notion of cardinality on which the natural numbers and the evens are of the same cardinality in virtue of the existence of a bijection, as opposed to a notion on which they are *not* of the same cardinality because one is a proper subset of the other. There are of course often good reasons for choosing one coherent precisification over another, reasons having to do with overall economy, tractability, fruitfulness, and so on. The important point here is that earlier practice in such cases does not pin down a collection of mathematical facts sufficient to dis-ambiguate the central terms. Instead, the reference-fixing is accomplished in part by straightforward decision: we stipulate that, in future, *this* is what we will mean by "continuous," by "infinite," by "of the same cardinality as," and so on. In these cases, one cannot say that later mathematicians expressed senses or referred to functions and objects that had been determinately pinned down by the combination of earlier usage and underlying mathematical facts.

Frege's own work towards conceptual clarification would seem to provide examples of just this kind. The reference of the numeral "2," as Frege presents it in *Grundgesetze*, is the extension of a first-level function (under which fall extensions of other first-level functions), while its reference as presented in *Grundlagen* is the extension of a second-level function (under which fall first-level concepts). Similarly for all other numerals. The change from the *Grundlagen* to the *Grundgesetze* numbers is not due to a change of view on Frege's part: the two treatments succeed for his purposes in just the same way. The change would seem to be driven by reasons of technical convenience. What this means, though, is that

the pre-Fregean use of the numerals, despite arguably fixing everything that matters from the point of view of pure arithmetic, does not fix enough to settle whether the numerals refer to the extensions of first-level functions. No arithmetical facts determine whether the ordinary 2 is the *Grundlagen*'s 2, the *Grundgesetze*'s 2, or something else altogether. Just as in the cases of ordinary mathematical development noted above, the reference of the terms in the mature version of the science is determined in part by *fiat*, and not just by plumbing the depths of those mathematical facts that have, all along, grounded the original practice.

Domain Expansion

The Fregean case just mentioned is arguably an example of a phenomenon that arises whenever we expand the domain of a mathematical theory. Having added extensions (or value-ranges) to the domain of discourse, we can frame new sentences, e.g. identity-sentences involving one term from the old theory and one from the expansion zone, whose truth-value is not fixed by anything that has gone before. No arithmetical facts determine whether zero is a value-range, and if so precisely which one. No facts about the cardinal numbers and the rationals determine whether the cardinal 2 is identical with the rational 2; similarly for expansions to reals and to complex numbers. This is of course as it should be: it makes no mathematical difference how we answer these "outlying" questions, and it would be absurd to expect that the domain of underlying mathematical facts has any bearing here. But thoughts are determinately true or false. If nothing about mathematical practice or about the facts to which we advert when carrying out that practice determines whether "$2_{card} = 2_{rat}$" expresses a truth or a falsehood, then nothing about that practice or about those facts determines what thought is expressed by that sentence. Similarly, nothing about that practice or about those facts determines whether two sentences differing just in the replacement of one such term for the other express the same thought, so that the indeterminacy would seem to affect not just such inconsequential sentences as "$2_{card} = 2_{rat}$," but virtually all sentences of the language.

The difficulty for Frege's view of the nature of mathematical discourse and of scientific continuity now seems to have deepened. Because later, cleaned-up versions of fundamental concepts often arise not just as the result of analyzing content, but in part as a result of making arbitrary decisions in the face of newly-recognized ambiguity, the idea that the original terminology had determinate reference seems to have been undermined. If there is no fact of the matter whether the ordinary "2" refers to a given extension, or whether early uses of "continuous" refer to Weierstrass's notion, then it would seem that these terms have no determinate reference—which is to say that they have no reference. The difficulty about thoughts is now not just the subtle question of whether one can make sense of thought-identity across significant conceptual development, but of whether one can make sense of the idea that ordinary mathematical discourse involves the expression of thoughts at all, in the face of this degree of ambiguity about reference.

Frege on Domain-Expansion

Frege recognizes two kinds of domain-expansion in mathematical theories: those in which the "added" objects are of a not strictly-mathematical kind, and so give rise to identity-statements linking e.g. numerical and non-numerical terms (for example, "2 = {{∅}}"), and those in which the "added" objects are from an enlarged but already-mathematical domain (e.g. "$2_{card} = 2_{rat}$"). In what follows, we examine his discussions of these cases, with an eye toward understanding to what extent the Fregean account of theoretical unity is undermined by domain-expansion. As we'll see, the difficulties for Frege are not negligible, but they are not as stark as has been suggested above.

In *Grundgesetze*, Frege introduces two kinds of singular terms: sentences (which, recall, are singular terms whose references are truth-values), and value-ranges. The truth-conditions of identity-sentences linking two value-range terms are given immediately by Law V, according to which the value-range of F = the value-range of G iff F and G give the same value for every argument. The truth-conditions of identity-sentences each of whose terms is a sentence are similarly straightforward: such identities are true if and only if the sentences on each side have (of course) the same reference, which is to say that they have the same truth-value. Left indeterminate by these factors, however, are the truth-conditions of identity-sentences in which the identity-sign is flanked by a sentence on one side, and a value-range term on the other. Such sentences will play no role in Frege's development of arithmetic, and hence, barring inconsistency, it does not matter how one fixes truth-conditions on them. But in keeping with Frege's insistence that every well-formed sentence of *Grundgesetze* have a determinate truth-value, it is essential that such sentences are fitted out with truth-conditions of some kind. Frege's way of meeting this requirement is simply to stipulate that all true sentences will refer to the value-range of any function under which exactly the True falls, and that every false sentence will refer to the value-range of any function under which exactly the False falls. The stipulation is arbitrary, in the sense that alternative stipulations could easily and unproblematically have been made in its place; the important point is simply that some coherent stipulation be made. Frege's remark about this stipulation is as follows:

> We have hereby determined the *value-ranges* as far as is possible here. Only when the further issue arises of introducing a function that is not completely reducible to the functions already known will we be able to stipulate what values it should have for value-ranges as arguments; and this can then be viewed as a determination of the value-ranges as well as of that function. [Frege (1893) §10)]

The interest of this passage is that it undermines what one might call a "naïve platonist" reading of Frege's understanding of the objects to which his singular terms refer. If we on a later occasion expand the language of *Grundgesetze* so as to make it suitable, say, for use in proofs about mechanics, we will introduce, amongst other things, new singular terms. The new "cross-category" identity sentences, i.e. those identifying a value-range and a new object, will have truth-conditions and

hence truth-values only after a further arbitrary stipulation is made, as described by Frege above. Hence there is no fact of the matter, prior to the stipulation, whether the terms in question co-refer. And this is not due to a failure of determinate reference on the part of the introduced terms; the indeterminacy of the identity-sentences obtains even when the newly-introduced terms are those of a fixed, determinate science, one whose claim to the expression of truth is as robust as possible.

The same holds not just for identity-sentences but, as Frege remarks above, for sentences that express the application of a function from the old theory to an object (or function, or n-tuple) from the new: prior to the imposition of some arbitrary stipulations, such sentences will frequently not have truth-values fixed either by the linguistic meanings of the terms, or by the underlying mathematical or other facts. That the stipulation needed in *Grundgesetze* §10 applies merely to identity-sentences is an artifact of the very simple language of that formal system.

The indeterminacy of cross-theory sentences is just what one should expect in the normal course of events: that the merging of two self-standing theories, or the simple expansion of a single theory, will give rise to cross-theory sentences whose truth-conditions aren't determined by any of the facts with which either theory (or: the original theory) is concerned is the standard case. But it's a situation that does not square well with a certain conception of what it is for the terms of the original theory or theories to have determinate reference. If one takes it that e.g. a function-term $f(x)$ and a singular term t both have determinate reference only if $f(t)$ does, and hence that singular terms t_1 and t_2 have determinate reference only if the identity-sentence $t_1 = t_2$ has a determinate truth-value, then the situation just described can only be understood as one in which the original theory or theories in question, no matter their long usefulness and success, have no terms with determinate reference. Given the possible (indeed, probable) expansion of mathematical terminology, and hence the possible (indeed probable) introduction into our vocabulary of novel cross-theory sentences of the kind just discussed, this conception is one on which none of our terms ever has determinate reference. So much the worse, of course, for the view that determinate reference in mathematics requires the kind of cross-theory determinacy just described.

It may seem, and indeed does seem to many, that Frege endorsed the platonic requirement on referentiality just discussed: the idea that determinacy of reference on the part of terms taken from different theories requires the determinacy of reference or truth-conditions for all syntactically-permissible combinations of those terms. The central reason one might have for attributing such a view to Frege is that one takes it to be an immediate consequence of his often-repeated claim that all functions are in some sense "total." But this, I take it, is a mistake. Frege's many and varied discussions of the requirement of totality for functions, which is to say the requirement that functions be defined for all arguments, are in every case discussions that apply to a single language: they are discussions of, indeed arguments for, the conclusion that rigor in formal systems requires that every function referred to in such a system is defined over every argument referred to in that system. This is, in short, the requirement of "linguistic completeness," the

requirement that every well-formed expression of a formal system has a determinate reference. I have argued elsewhere, so won't go into details here, that Frege's commitment to the requirement of linguistic completeness is absolute for formal languages, that he holds no such requirement for languages of ordinary discourse, and that he does not hold the considerably stronger requirement that the functions referred to in a given system be defined over arguments from outlying areas.[4] Frege is not, in short, a platonist in the above sense about reference.

Regarding the broadening of the sense and reference of such terms as "function" and "sum," Frege remarks as follows:

> Now how has the meaning of the word 'function' been extended by the progress of science? We can distinguish two directions in which this has happened.
>
> In the first place, the field of mathematical operations that serve for constructing functions has been extended. Besides addition, multiplication, exponentiation, and their converses, the various means of transition to the limit have been introduced…
>
> Secondly, the field of possible arguments and values for functions has been extended by the admission of complex numbers. In conjunction with this, the sense of the expressions 'sum,' 'product,' etc. had to be defined more widely. [Frege (1891) 12, (1984) 144]

Specifically regarding the addition-function, we find:

> After thus extending the field of things that may be taken as arguments, we must get more exact specifications as to what is meant by the signs already in use. So long as the objects dealt with in arithmetic are the integers, the plus-sign need be defined only between integers. Every widening of the field to which the objects indicated by *a* and *b* belong obliges us to give a new definition of the plus-sign. [Frege (1891) 19, (1984) 148]

Frege has, in short, the ordinary mathematician's view about the development of a given theory into a new domain: that the widening of the objects dealt with requires a widening of the domains of the relevant functions, but that this ever-present possibility is no hindrance to perfectly determinate reference on the part of the original terms in their old settings.

The widening of the domain of a function is strictly speaking a matter of dealing with a new function; as a consequence, Frege takes it that strict logical rigor mandates in such cases the use of a new term. As he puts it in the second volume of *Grundgesetze*,

> *§56.* A definition of a concept (a possible predicate) must be complete; it has to determine unambiguously for every object whether it falls under the concept or not…
>
> *§57.* From this now follows the inadmissibility of piecemeal definition, which is so popular in mathematics. This consists in providing a definition for a special case – for example, for the positive whole numbers – and putting it to use and then, after various theorems, following it up with a second explanation for a different case – for example, for the negative whole numbers and for Zero – at which point, all too often, the mistake is committed of once again making determinations for the case already dealt with. …
>
> *§58.* To be sure, we have to grant that the development of the science which occurred in the conquest of ever wider domains of numbers almost inevitably demands such a practice; and this demand could be used as an apology. Indeed, it would be possible to replace the

[4] See Blanchette (2012a, b).

> old signs and notations by new ones, and actually, this is what logic requires; but this is a decision that is hard to make. …
>
> *§60:* It is, moreover, very easy to avoid multiple explanations of the same sign. Instead of first explaining it for a restricted domain and then using it to explain itself for a wider domain, that is, instead of employing the same sign twice over, one need only choose different signs and to confine the reference of the first to the restricted domain once and for all, so that the first definition is now also complete and draws sharp boundaries. [Frege (1903) §§56–60]

The earlier sign, with restricted domain, has unproblematic reference despite remaining undefined over objects that lie outside the bounds of its theory. Frege makes the same point in his lectures of the summer of 1914, if Carnap's notes are accurate:

> In the development of mathematics one does, however reach certain points where one wants to expand the system. But then one has to begin from scratch again. In any case, there always has to be a complete *system* at hand that is logically unproblematic. E.g. one would have to proceed as follows: as long as the plus sign + is used only for positive whole numbers, one chooses a different sign for it, e.g., ¬. [Reck and Awodey (2004), p. 155]

In short: the difficulty most recently mentioned, i.e. that Frege's requirement of total definition for functions makes impossible the recognition of cross-theory sentences whose terms each have determinate reference in their original setting, but whose own truth-conditions are settled only by stipulation, is ill-founded. Frege is in this sense a perfectly ordinary mathematician, one who takes it that the new sentences yielded by an expansion of the domain of a mathematical theory will include some whose truth-value is determined by the mathematical facts, and some whose truth-value can be fixed only by arbitrary stipulation.

Reference

The platonist conception of reference, to which we have contrasted Frege's, is a conception on which the determinacy of reference in their own distinct settings of a function-term $f(x)$ and a singular term t requires that the cross-theoretical sentence $f(t)$ have a determinate truth-value. As we have seen, this conception is neither plausible as a constraint on reference in mathematics, nor plausibly attributed to Frege. But to say that Frege is not a platonist about reference is not to say that he lacks stringent requirements on referential terms within rigorous theories. We turn here to a brief account of Frege's requirements on reference in *Grundgesetze*.

In *Grundgesetze* I §29, Frege gives the following sufficient conditions for reference:

- A one-place first-level function-name has a reference if every result of filling its argument-place with a referring proper name has a reference.
- A proper name has a reference if the result of using it to fill an argument-place of a referring first-level function-name itself always has reference.
- And so on.

Taking for granted that some simple sentences express truths or falsehoods, sections 30–32 contain a rigorous proof that all well-formed names (including sentences) of *Grundgesetze* have determinate reference.

In these sections of *Grundgesetze*, we get a clear picture of exactly what, according to Frege, is required in order for a piece of language in a mathematical theory to count as having determinate reference. The requirement is very strictly theory-bound: what's required is that, as proven in §§30–32, each function-term provides a determinate value when given as argument any object in the domain of the theory (which in *Grundgesetze* includes just the two truth-values and value-ranges of first-level functions; recall that numbers are value-ranges). The clear and rigorously-demonstrated view that every well-formed piece of *Grundgesetze* notation has a determinate reference is not undermined by the similarly clear, and clearly-acknowledged, fact that the functions in question are not defined over "outlying" objects. The referentiality of a term *t* of *Grundgesetze* by no means requires that identity-sentences linking *t* with terms from outside the theory have truth-values or truth-conditions. Finally, it is worth recalling that, as Frege understands it, the well-formed terms of *Grundgesetze*, including its sentences, have been shown by the reasoning in §§ 30–32 not just to have determinate reference, but to have sense as well.[5]

Frege's fundamental picture of reference as it applies to a mathematical theory is that a mathematical term has sense and reference if our understanding of that term (supplemented if necessary by stipulations), together with mathematical facts, fix the truth-values of sentences formulable in the theory. That the terms of a theory have determinate sense and reference is compatible with two kinds of ignorance on the part of its users. It is compatible with intratheoretical ignorance, i.e. ignorance of answers to questions formulable in the theory, as long as those answers are determined by mathematical facts. We can, for example, use the terms of number theory entirely competently while remaining in ignorance of the truth-value of Goldbach's conjecture. The second kind of ignorance compatible with the competent use of the theory is extratheoretical, i.e. ignorance of answers to questions not formulated in the theory as it stands. Indeed, it is no part of Frege's theory to suppose that the latter kinds of question have answers at all.

Thought-Identity

As argued above, Frege's conception of sense and reference is one on which the kinds of domain-expansion that go along with scientific progress are compatible with the possession of determinate sense and reference by the terms of early-stage

[5]"Thus it is shown that our eight primitive names have a reference and thereby that the same applies to all names correctly formed out of them. However, not only a reference but also a sense belongs to all names correctly formed from our signs. Every such name of a truth-value *expresses* a sense, a *thought*." [Frege (1893) §32].

theories. The pressing question, now, is whether scientific continuity in the face of progress is compatible with Frege's idea of continuity as involving the preservation of thought.

To get an idea of Frege's understanding of the sense in which theoretical continuity is possible in the face of conceptual progress, we will look at his own project, that of providing a newly-rigorous and clarified version of arithmetic in *Grundlagen* and *Grundgesetze*. We should note at the outset that the move from arithmetic as ordinarily understood to Frege's rigorous new account of it involves both of the kinds of development discussed above: the new theory has a broader domain (including, in the case of *Grundgesetze*, infinitely many value-ranges), and its sentences are not easily-recognizable synonyms of their original counterparts.

We begin with *Grundlagen*. The central question is this: what, exactly, does Frege take his well-developed theory of *Grundlagen* to "preserve" with respect to ordinary arithmetic? We note first that Frege is centrally concerned in *Grundlagen* with biconditionals of the following forms:

(i) The number of F's = the number of G's iff there's a bijection of the F's onto the G's;
(ii) The number of F's = 0 iff $\forall x \sim Fx$;
(iii) The number of F's = n + 1 iff $\exists a$ (Fa & the number of (F's other than a) = n).

Much of *Grundlagen* is taken up with arguing that the left-hand side of (i), as ordinarily understood, expresses essentially what its right-hand side expresses. Similarly for (ii) and (iii). Frege's suggestion is that in each case, the right-hand side provides an adequate analysis of the ordinary content found on the left.[6]

If one grants Frege this non-trivial claim of analytic adequacy, the rest of the *Grundlagen* project flows smoothly (or would have, if not for the difficulty about extensions). Terms of the form "the number of F's" refer, in the theory of *Grundlagen*, to extensions of concepts in such a way that the left-hand sides of (i)–(iii), so understood, are immediately logically equivalent with the respective right-hand sides of the biconditionals, as ordinarily understood.[7] This means that the connection between statements of the form "The number of F's = the number of G's" as ordinarily understood, and those statements as understood via the new *Grundlagen* account, is straightforward: in each case, the latter is logically equivalent with a good analysis of the former. Similarly for (ii) and (iii). If we focus just on sentences of the form (i)–(iii), what is "preserved" in the move from the early theory of ordinary arithmetic to its development as *Grundlagen* arithmetic is quite significant: each sentence of the latter camp is obviously logically equivalent with its counterpart from the former. Because Frege develops the whole of the arithmetic of the finite cardinals from these sentences, the result is that the arithmetical claims expressible in the new theory have the same grounds as do claims expressible in the

[6]This account of Frege's project as one of conceptual analysis is argued for in Blanchette (2012a).

[7]The logical equivalence here requires the faulty principle about extensions, assumed by Frege in *Grundlagen*, that the extension of F = the extension of G iff $\forall x$ (Fx iff Gx).

old: if Frege's project of proving *Grundlagen* arithmetic from principles of logic had succeeded, he would arguably, modulo agreement about his original analysis, have succeeded in demonstrating that the claims of *ordinary* arithmetic are, themselves, grounded in pure logic.

This picture of the relationship between the earlier theory, that of ordinary arithmetic, and its later development in a more sophisticated framework is exhibited again in Frege's mature development of the theory in *Grundgesetze*. Here too we find an account of terms of the form "the number of F's", and of "0" and "successor" that suffice to deliver the result that: if all had gone well, the thoughts expressed in the new theory, thoughts about value-ranges of functions, would have been clearly logically equivalent with good analyses of their ordinary counterparts. There is no question, in Frege's development, of trying to assign to the numerals of the new theory, taken in isolation, the same sense or reference as is had by the numerals of ordinary arithmetic. The question of whether the "2" of *Grundgesetze* co-refers with the "2" of ordinary arithmetic is, as above, not a well-formed question from the Fregean point of view. But what *is* preserved is what matters for a foundational project: had all gone well with the value-ranges, the truths of the newly-constructed arithmetic would have been self-evident logical equivalents of their ordinary counterparts.

In sum: Frege's strategy in both *Grundlagen* and *Grundgesetze* is to provide an accurate account of central notions—e.g. "cardinal number," "0," "successor," "finite cardinal" by providing accurate accounts of the core statements of the theory of these notions, including centrally our (i)–(iii) above. The accuracy of the accounts is judged via the question of whether these core sentences and their analysantia are sufficiently similar, semantically, that a proof of the latter from a given collection of premises will suffice to establish the logical entailment of the former by those premises. The success of *Grundlagen* and *Grundgesetze* in this attempt was to have turned on (a) the success of the original analyses, as discussed above in connection with biconditionals (i)–(iii), and (b) the correctness of Frege's assumption that Law V (or its counterpart with respect to extensions) was a law of logic.

Do the sentences of *Grundgesetze* express the same thoughts as do their counterparts in ordinary arithmetic? Here, the only thing to say is that there is no clear answer. Frege's guiding ideas that sentences express thoughts, and that thoughts are the constituents of theories, do not come along with rigorous criteria of thought-identity. To answer our question, one would have to know whether in Frege's view a successful analysis, of the kind linking "the number of F's = the number of G's" and "there is a bijection of the F's onto the G's" delivers pairs of sentences that express the same thought. To this question, Frege simply gives no answer. But perhaps most relevant to the questions raised at the outset of this essay regarding the consistency of Frege's view of theories with the phenomena of mathematical progress are the following points that have now become clear. First of all, the fact that the sentences of *Grundgesetze* differ in cognitive significance from those of ordinary arithmetic is no reason to conclude that the thoughts expressed thereby are different. Similarly, the fact that "cross-theory" identity statements,

those linking terms of ordinary arithmetic with terms of *Grundgesetze* notation, lack truth-conditions is no barrier to the determinacy of reference and sense on the part of those terms. Neither, finally, is the indeterminacy of cross-theory identities a barrier to the idea that corresponding sentences of the two theories express the same thought. That no aspect of the ordinary use of "the number of even primes" determines that this term in its ordinary use co-refers with any term of *Grundgesetze* is not by itself a barrier to the expression of the same thought by the ordinary "the number of even primes = the number of positive square roots of 9" and its *Grundgesetze* counterpart. The question of whether they do express the same thought turns on the questions of whether good conceptual analysis preserves thought, and of whether the two sides of an instance of Law V express the same thought.

Perhaps most importantly: the theoretical continuity essential for Frege's project is not, in the end, that sentences from the old and the new theories express the *same* thought, but that the thoughts they express are sufficiently similar for the purposes of the logicist project. Essential here are the fact that new and old counterparts have the same truth-conditions, and even more robustly that any premises providing a logical ground of one will suffice to ground the other. Whether the relation between ordinary and *Grundgesetze* sentence is understood as thought-identity, in line with a very coarse-grained account of the identity-conditions of thoughts, or instead as a broader kind of content-similarity, in line with a more fine-grained such account, is not determined by Frege's general views about mathematics or about thoughts. This is, one might say, as it should be, since it makes no difference to Frege's project or to his understanding of mathematics how one decides to describe the situation.

Similarly for mathematical progress generally. The move from earlier theories whose concepts are relatively vague to those with more sharply-defined notions and perhaps with larger domains is one that requires a certain recognizable similarity between the thoughts expressed, and not merely, for example, mere similarity in syntactic or proof-theoretic structure. Frege's work gives no algorithm for the precise similarity necessary for theoretical continuity, just as it gives no precise criterion of thought-identity. The former is presumably what one should expect, since the precise kind and degree of similarity required for continuity will vary from field to field and era to era. Frege therefore simplifies and over-states the case when describing theoretical continuity in terms of thought-identity. But from what we have seen so far, his fundamental conception of mathematical progress and continuity is not threatened by the phenomena that at first seemed problematic: the non-trivial conceptual clarification and the domain expansions that go along with mathematical progress.

Fleshing Out the Positive Picture

We turn, finally, to one particular aspect of Frege's conception of theoretical continuity in mathematics that helps to fill out his positive view of the similarity required between earlier and later theories. As a description of his own project, Frege says that:

> [F]or every object there is one type of proposition which must have a sense, namely the recognition-statement, which in the case of numbers is called an identity. Statements of number too are, we saw, to be considered as identities. The problem, therefore, was to fix the sense of a numerical identity … [Frege (1884) §106]

Similarly,

> [F]rom our previous treatment of the positive whole numbers, [we] have seen that it is possible to avoid all importation of external things and geometrical intuitions into arithmetic, without, for all that, falling into the error of the formalists. Here, just as there, it is a matter of fixing the content of a recognition-judgment. Once suppose this everywhere accomplished, and numbers of every kind, whether negative, fractional, irrational, or complex, are revealed as no more mysterious than the positive whole numbers … [Frege (1884) §109; emphasis added]

And with respect to later expansions to larger classes of numbers, we find again the same fundamental ideas: that the analysis of the theory of those numbers is to turn on the analysis of a collection of core sentences, and that those core sentences are identity-sentences involving the numbers in question:

> How are complex numbers to be given to us then, and fractions and irrational numbers? If we turn for assistance to intuition, we import something foreign into arithmetic ….
>
> [*review of Grundlagen's account of Number ... - PB*]
>
> In the same way with the definitions of fractions, complex numbers and the rest, everything will in the end come down to the search for a judgment-content which can be transformed into an identity whose sides precisely are the new numbers. [Frege (1884) §104]

It is worth noting that if, as Frege holds, the core to be preserved in the development of a theory includes (or, as above, is composed of) identity sentences, then it is essential that the content of the identity sign not be re-defined as we move from theory to theory. This is not a trivial point, and would not have been uncontroversial to Frege's readers: it was, indeed still is, common to take a certain form of definition as involving the redefinition of identity over a given domain: consider for example the "identification" of diametrically opposed points in a spherical model of geometry, the "identification" of multiple pairs of integers as the same rational, and so on. Because Frege's understanding of what's preserved across developments of a given subject-matter includes the content of core identity-sentences, it is essential (in order that this requirement be non-trivial) that the identity-sign in question expresses the real identity-relation, and not some simulacrum, across all of the theories in question. We see this requirement in operation in Frege's response to Peano. First, the relevant passage from Peano, as quoted by Frege:

> [G]iven equality between integers, one defines equality between rationals, between imaginary numbers, etc. In geometry one is used to defining the equality of two areas, of two volumes, the equality of two vectors, etc. … The various references have properties in common; but I do not see how they suffice to specify all the possible references of equality. … Moreover, the opinions of the various authors concerning the concept of equality are very diverse…
>
> [Peano (1898), as quoted by Frege in Frege (1903) §58 n]

Frege's predictable reply is as follows:

> If mathematicians' opinions about equality diverge, then this means nothing less than that mathematicians disagree with respect to the content of their science; and if one views the essence of the science as being thoughts, rather than words or signs, then this means that there is no one united mathematical science, that mathematicians do not, in fact, understand each other. For the sense of nearly all arithmetical propositions and of many geometrical propositions depends, directly or indirectly, on the sense of the word 'equal.' [Frege (1903) §58 n]

This returns us essentially to our starting-point. Two mathematicians are engaged in the same science only if each of them expresses thoughts that are appropriately related to the other's. For a science is simply a body of thoughts.

The idea that theoretical continuity turns on the successful treatment of a handful of core sentences means that, from the Fregean perspective, the question of continuity is at least in some cases a relatively tractable one: a proposed development of a theory counts as merely changing the subject if it fails to get the core sentences right—i.e. if it assigns to those sentences thoughts insufficiently similar to the thoughts they ordinarily express. This is the criticism leveled by Frege in *Grundlagen* at those accounts of arithmetic that construe the core sentences as expressing claims about geometry or about ideas, and it is his perennial criticism of the so-called "formalist" accounts of arithmetic.

Once we get the core sentences right, Frege seems to say, we count as thereby having remained faithful to the central concepts of the science. Amongst the general requirements we have seen, above, for doing so is the further requirement that the developed theory maintains the original relation of identity: no relation that, for example, holds between distinct objects can be taken as the reference of the identity-sign, at the risk of entirely undermining the required similarity between the core identity sentences and their developed counterparts.

Conclusion

We have seen that Frege's requirements for theoretical continuity are not what one might have taken them to be. Theoretical continuity does not require preservation of conventional significance. It also does not require preservation of reference at the level of individual terms. It requires instead a certain imprecise but often recognizable similarity of thought expressed by each of the pairs of core sentences. Whether this relationship between the sentences is to be understood as the

expression of the *same* thought, as opposed to that of recognizably-similar thoughts, has turned out to be unimportant, neither entailed nor contradicted by Frege's central views. Finally, Frege's guiding principle, according to which theoretical continuity requires the kind of thought-similarity discussed above, though less than precise, is a forceful view, ruling out for example all of those attempts to develop arithmetic on geometric, formalist or idealist grounds, at least without supplementary argument to the effect that the core sentences of arithmetic have, as ordinarily understood, a geometric, formalist or idealist content.

References

Blanchette, P. (2012a). *Frege's conception of logic*. New York: Oxford University Press.

Blanchette, P. (2012b). Frege on shared belief and total functions. *The Journal of Philosophy CIX, 1/2*, 9–39.

Burge, T. (1979). Sinning against frege. *The Philosophical Review*, *88*, 398–432. (Reprinted in Burge (2005) 213–239).

Burge, T. (1984). Frege on extensions of concepts, from 1884 to 1903. *The Philosophical Review*, *93*, 3–34. Reprinted in Burge (2005) 273–298.

Burge, T. (2005). *Truth, thought, reason: Essays on Frege*. Oxford: Clarendon Press.

Frege, G. (1884). *Die Grundlagen der Arithmetik.* Breslau: William Koebner. [English edition: Frege, G (1978). *The foundations of arithmetic*]. (J. L. Austin, Trans.). Evanston, Ill.: Northwestern University Press.

Frege, G. (1891). *Funktion und Begriff*. Jena: Hermann Pohle. Reprinted in Frege (1967): 125–142. [English edition: Frege, G. (1984a). Function and concept. Frege (1984): 137–156] (P. Geach, Trans.).

Frege, G. (1892a). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*, *100*, 25–50. Reprinted in Frege (1967): 143–162. [English edition: Frege, G. (1892). On Sense and Reference. Frege (1984): 157–177] (M. Black, Trans.).

Frege, G. (1892b). Über Begriff und Gegenstand. *Vierteljahrsschrift für wissenschaftliche Philosophie*, *16*, 192–205. Reprinted in Frege, G. (1967): 167–178. [English edition: Frege, G. (1984b). On Concept and Object. Frege (1984): 182–194] (P. Geach, Trans.).

Frege, G. (1893). *Grundgesetze der Arithmetik* (Vol. I). Jena: Hermann Pohle. [English edition: Frege, G. (2013). *Basic laws of arithmetic* (P. Ebert, & M. Rossberg, Trans.). Oxford: Oxford University Press.

Frege, G. (1903). *Grundgesetze der Arithmetik* (Vol. II). Jena: Hermann Pohle. [English edition: Frege, G. (2013). *Basic laws of arithmetic* (P. Ebert, & M. Rossberg, Trans.). Oxford: Oxford University Press.

Frege, G. (1967). *Kleine Schriften*. In I. Angelelli (Ed.), Olms: Hildesheim.

Frege, G. (1979). *Posthumous writings*. In H. Hermes, F. Kambartel, & F. Kaulbach (Ed.), (Long & White, Trans.). Chicago: University of Chicago Press.

Frege, G. (1983). *Nachgelassene Schriften*. In H. Hermes, Friedrich Kambartel, & F. Kaulbach (Eds.) Hamburg: Felix Meiner Verlag.

Frege, G. (1984). *Collected papers on mathematics, logic, and philosophy*. B. McGuinness (Ed.), Oxford: Blackwell.

Peano, G. (1898). Riposta. *Rivista di Matematica, 6*, 60–61.

Reck, E., & Awodey, S. (Eds.). (2004). *Frege's lectures on logic: Carnap's student notes, 1910–1914*. Chicago: Open Court.

Author Biography

Patricia Blanchette is Professor of Philosophy at the University of Notre Dame. She is the author of *Frege's Conception of Logic* (Oxford University Press 2012), and of numerous articles on Frege and on the history and philosophy of logic.

Identity in Frege's Shadow

Jaakko Hintikka

Frege's Mistake

One of the crucial figures in early analytic philosophy was Gottlob Frege. This is not due to his direct influence. Frege exerted his influence through his creation, his logic. (The fullest exposition of this logic is in Frege 1893–1903a see also 1893–1903b.) This logic has been used without major changes by the majority of subsequent philosophers. As a consequence its strengths and the weaknesses have been magnified by history. This essay is written not to praise Frege but to bury—or at least to diagnose—one particularly important mistake affecting Frege's logic. The later effects of this virus in Frege's logic are too exstensive to be discussed here. (See here especially Hintikka, forthcoming).

Frege is usually credited with having created the core part of modern symbolic logic, the theory of quantifiers which is essentially what we now call first-order logic. (Other names that have been used include predicate calculus, lower functional calculus and quantification theory). This may be true in some historical sense, but if so, Frege did not finish his task. He made a mistake. He did not fully understand the semantical job description of quantifiers and as a consequence produced a logic that is unnecessarily restricted in its expressive capacities. He thought that the job of quantifiers is exhaustively done by their ranging over a class of values, so as to express exceptionlessness or nonemptiness in this class. He expressed this assumption in so many words in that he characterized quantifiers as higher-order predicates indicating emptiness or exhaustiveness of a given

Jaakko Hintikka—deceased

J. Hintikka (✉)
Boston University, Boston, USA
e-mail: hintikka@bu.edu

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_2

lower-order predicate (Frege 1893–1903b). The full range of values of quantified first-order variables is in our days taken to be the domain of a model alias its universe of discourse.

This "ranging over" is admittedly a part of the semantical story of quantification. But there is also a subtler and more interesting task that quantifiers perform. This other task is vitally important in mathematics, science and in everyday life. It is to express dependence relations, that is, dependencies (and by contrast independencies) between variables. On the first-order level of a traditional logical language, the only way of expressing that y depends on x is to make the quantifier (Q_2y) to which y is bound formally dependent (or formally independent) of the quantifier (Q_1x) to which x is bound.

The simplest questions (but not the only ones) that bring out this other function of quantifiers concern quantifier ordering, including the effects of changes in such ordering. "Every man loves some woman" is not equivalent with "Some woman is loved by every man" because in the latter the woman depends on the lover while in the former she does not. Yet the only formal difference between them lies in the (logical) order of the two quantifiers "some" and "every". This order is independent of the range of the quantifiers involved. It matters even if we are conceptualizing the situation in terms of many-sorted logic in which "every man" and "some woman" range over different classes of persons. Questions of quantifier ordering thus form a handy testing ground for problems of quantifier dependence (and independence).

In the received ("Fregean") first-order logic (alias RFO logic), formal dependencies among quantifiers are expressed by the scope brackets (). A quantifier (Q_2y) depends on (Q_1x) if it occurs within its scope, as in

$$(Q_1x)(\ldots(Q_2y)(\ldots)\ldots) \tag{1.1}$$

Scopes are assumed to be nested. There is in RFO logic no such thing as (partial) overlap of scopes. As a result, the dependence structures expressible in RFO logic are finite labeled trees, with dependence relations that exhibit a partial ordering which is linear in one direction. Such dependence structures are nevertheless only a proper subclass of all possible dependence structures. The rest cannot be expressed in RFO logic, which is therefore not a full account of the semantics of quantifiers. Frege's logic, and the logics used by most subsequent logicians and philosophers, are therefore defective. Their expressive power is restricted unnecessarily, seriously handicapping them.

If Logic Corrects Frege's Mistake

This mistake of overlooking the dependence-indicating function of quantifiers will be called in this paper for brevity Frege's Fallacy. It was not unavoidable, and it was not unavoidable in Frege's actual historical situation. Already at Frege's time Charles S. Peirce understood the dependence-indicating aspect of first-order logic

better than Frege. The term "Frege's Fallacy" may nevertheless be somewhat unfair in that Frege's failure can be seen as a sin of omission rather than as one of commission.

Frege's mistake can be corrected (see here Hintikka 1996). Systematically the minimal improvement is to introduce a slash notation which makes an existential quantifier (∃y) independent of a universal quantifier (∀x) (in a sentence in a negation normal form) within whose formal scope it occurs by writing it as (∃y/∀x). This results in what is known as independence friendly (IF) logic. This logic is thus not a "nonclassical" variant or a special branch of RFO. It is not an "alternative logic". It replaces RFO by correcting its shortcomings. It is as classical or non-classical as RFO. It contains RFO as a special case, viz. as the logic of those propositions that satisfy the tertium non datur. The status of IF logic as being nothing more and nothing less than an improved version of RFO logic is illustrated by the fact that it could be formulated without any new symbols merely by liberalizing the use of parentheses.

The replacement of RFO by IF first-order logic is nevertheless more than a mere architectonic improvement. Frege's mistake has been recommitted by most subsequent logicians, including the leading ones. It has led to damaging missteps in the development of logic and its applications. On other occasions I will show that Frege's Fallacy has played a role in such important developments as Frege's higher-order logic and Hilbert's epsilon-calculus. It was responsible for the main paradoxes of set theory and hence indirectly for the invention of the theory of types. Less directly, Frege's mistake has played a major role in the (not so new) New Theory of Reference due to Kripke (1963). Here, only some general explanation explanations are offered.

Skolem Functions as Dependence Indicators

The dependence indicating task of quantifiers deserves closer examination. In the abstract eyes of a mathematician, to speak of dependencies is to speak of functions, that is, of the functions that codify those dependencies. Hence, the logic of quantifiers qua dependence indicators is a logic of functions. The functions that are tacitly present in quantificational propositions are the ones that express the dependencies between quantifiers, that is, express how the truth making value of a dependent (embedded) quantified variable is determined by the values of the quantifiers in whose scope it occurs. The functions that serve this purpose in a given quantificational sentence S are in logic called the Skolem functions of S. For instance, the Skolem function *f* of a sentence of the form

$$(\forall x)(\exists y)\mathrm{F}[x, y] \tag{3.1}$$

makes true the sentence

$$(\forall x)\mathrm{F}[\mathrm{x}, f(\mathrm{x})]. \tag{3.2}$$

Hence the real logic of quantifiers, the real first-order logic is in effect the logic of Skolem functions, not a "predicate logic" (as it is often called). This might seem to be a mere terminological nuance, but in reality it is a symptom of a deeper difference.

The nature of quantification theory as being virtually but a theory of Skolem functions is highlighted by the intimate connection between Skolem functions and the truth conditions of quantified propositions. The most natural truth predicate of first-order sentences is in fact the existence of (a full set of) their Skolem functions.

Because of this role of Skolem functions, we have to consider first-order logic of functions as being separate from predicate logic. The two have different logical truths, it turns out.

This important fact is not generally recognized. Most logicians treat functions as predicates (relations) of a certain kind. More explicitly expressed, an equation like $\mathrm{y} = f(\mathrm{x})$ is treated as if it were a binary relation F[x, y], with the following special properties

$$(\forall x)(\exists \mathrm{y})\mathrm{F}[\mathrm{x}, \mathrm{y}] \tag{3.3}$$

$$(\forall x)(\forall y)(\forall z)\left((\mathrm{F}[\mathrm{x}, \mathrm{y}] \,\&\, \mathrm{F}[\mathrm{x}, \mathrm{z}]) \supset \mathrm{y} = \mathrm{z}\right). \tag{3.4}$$

The corresponding function is then defined in terms of F by the equivalence

$$(\forall x)(\forall y)\left((\mathrm{F}[\mathrm{x}, \mathrm{y}] \leftrightarrow \mathrm{y} = f(\mathrm{x})\right). \tag{3.5}$$

But (3.3) and (3.4) are not logical truths about the predicate F. For each function f they have to be introduced as ad hoc premises. If f is a function, the assumptions (3.3) and (3.4) should be logical and not merely contingent truths. Hence the logic of functions and the logic of predicates have different logical truths.

The important differences between the logic of functions and the logic of predication are not always appreciated. They can nevertheless be important here. Perhaps the most consequential difference is that even if functions are treated as predicates they cannot be adequately defined, in the sense that the definitions could be treated as conceptual (logical) truths. And if individuals are construed as constant functions, by the same token they cannot be logically defined, either.

This straightforward logical observation has significant implications. For instance, it makes a significant difference to the logical theory of Frege and Wittgenstein (the author of the *Tractatus*) that they did not have functions among their nonlogical primitive constructs (cf. Hintikka 2004).

Quantifiers and Identity

In the dichotomy between the logic of functions and the logic of predication, the notion of identity belongs to the logic of functions and quantifiers. The truth conditions (whatever they may be) of a predication F[a, b] normally turns on what a and b are like on their own or in relation to each other. In contrast, the truth conditions of an equation $y = f(x)$ depend in an obvious sense only on the identity of x and y, that is on which entities x and y are. This is because the function so to speak enables one to find y (whatever object that is) on the basis of merely knowing what x is.

The concrete meaning of this informal observation is illustrated by the fact that quantifiers depend on the mode of identification of the values of the variables. Without trying to spell out the theory of identification here (I have done it elsewhere) suffice it to say as an example that although perspectivally identified and publicly identified objects are not two separated classes of objects existing side by side, they are values of two different pairs of quantifiers.

The same connection between quantifiers and the notion of identity is manifested also in the fact that the substitutivity of identicals must hold for the variables of quantification, other things being equal. The situation is nevertheless complicated by the fact that other things are not always equal. There are different modes of identification, and even apart from them there are different meanings of identity. Comments on these problems can be found in my earlier papers.

Ironically, this feature of the logic of functions may have encouraged Frege's Fallacy. Since what matters in functional equations is apparently only the identity of the terms involved, it might seem that what matters in interpreting quantifiers is merely the class of their possible values.

Substitutional Interpretation of Quantifiers

To return to Frege's Fallacy, most philosophers who commit it do so tacitly, in the privacy of their presuppositions. There is nevertheless a way of committing oneself to the fallacious view publicly. It is to adopt what is usually referred to as the substitutional interpretation of quantifiers. The gist of this so-called interpretation is often taken to lie in formulating truth-conditions for quantificational sentences in terms of names and substitution-instances of formulas using names, and contrasted to what is known as objectual interpretation. However, this is a distinction without much structural difference, as is argued passionately by Kripke (1963). He is right. Whatever differences there may between the two so-called interpretations, they do not need to affect any formal (syntactical) laws or principles. In this sense, the difference lies only in a different way of speaking of the truth-conditions of quantificational sentences. Vienna Circle members might have spoken here of material and formal modes of speech.

In other words, the contrast between the two so-called interpretations is often discussed by asking such questions as to whether there are enough names to capture substitutionally the truth-conditions of quantificational sentences in infinite and perhaps even uncountable domains. Questions like these nevertheless do not cut very deeply. Mathematicians operate routinely with infinite and even uncountable sets. What is so strange about infinite sets of names? Are not ordinary numerals such a set? Furthermore, those names need not be actually present in one's object language. For substitutional interpretation it suffices that we talk about them in a suitable metalanguage.

Kripke [8] has seen correctly where the conceptual gist of the substitutional interpretation lies. It lies in the claim that the truth of a quantificational sentence equals the truth of a certain (possibly infinite) truth-function of atomic sentences and/or of their negations. (For a qualification that does not matter here, see below.) For instance, a universally quantified sentence $(\forall x)F[x]$ is logically equivalent with the conjunction of its substitution-instances, i.e. equivalent with

$$F[a_1] \,\&\, F[a_2] \,\&\ldots\&\, F[a_i] \,\&\ldots \tag{5.1}$$

where the a_i are all the members of the universe of discourse.

But to maintain this is to take the semantical function of quantifiers to be exhausted by their ranging over their values, in this case over the set $\{a_i\}$. In other words, a strictly understood substitutional interpretation is but another form of Frege's Fallacy.

The Inconspicuousness of Frege's Fallacy

The Fregean interpretation of quantifiers has in fact been adopted by some philosophers, for example by the author of *Tractatus Logico-Philosophicus*. Alas, Wittgenstein later came to consider his treatment of quantifiers the "biggest mistake" he made in the Tractatus.

What is wrong with the substitutional interpretation? What are symptoms of the mistakes that Frege's Fallacy causes? One might perhaps expect the problems to show up in the dependence relations between ordinary quantifiers, for instance in the behavior of quantifier ordering. As it happens, such problems are nevertheless all too easily hidden—or, more accurately, brushed under a carpet. The carpet is the usual interpretation of propositional logic. In the substitutional interpretation dependence relations between quantifiers become dependence relations between propositional connectives, which are far less conspicuous than the corresponding relations between quantifiers This does not change the problem, however. The possible patterns of dependencies are analogous in the two cases, and equally in need of liberation. However, the difficulties are much less clearly in evidence in the case of propositional connectives than in the case of quantifiers. Dependence and independence relations between connectives have not attracted much attention by

logicans. This is not to say that there is no need of enriching propositional logic, too, by freeing it from the fetters of the conventional scope relation. Indeed, computer scientists are beginning to make use of such liberated patterns of propositional dependence relations. These different patterns are in fact relevant to questions of computer architecture.

Frege's Fallacy therefore rears its ugly head most obviously with quantifiers other than the two standard quantifiers of first-order logic. A prime example is offered by the quantificational character of modal and epistemic notions. In using modal logic we are in effect quantifying over (and into) *possibilia* of some kind or other. For instance the necessity operator employed in possible-world semantics functions in effect as a universal quantifier ranging over possible worlds. One does not have to be Leibniz or Carnap to agree that a sentence is necessary true if and only if it is true in all possible worlds. Likewise, the sentence "it is known that *S*" (in brief, K*S*) is true if and only if *S* is the case in all scenarios compatible with what is known. Hence quantifiers and modal operators can be expected to depend on each other, unless specific assumptions to the contrary are made. One such assumption is the substitutional interpretation of quantifiers. It implies that quantifiers are independent of modal operators in that they range over the same class of values irrespective of the (modal) context. Since that class is the universe of discourse and since modal operators range over possible worlds, it follows that all different possible worlds must have the same domain of individuals (the same universe of discourse).

Logicians' attention has been distracted from the interplay of such quasi-quantifiers with regular quantifiers by their being of a different logical type and hence ranging over different kinds of entities. However, as was pointed out, differences in range do not make any difference to questions of dependence.

Quantifiers Depend on Modal Operators

What creates a new situation is that in modal and intensional logic the dependence relations that various quantifiers serve to express cannot any longer be hidden behind the back of dependence relations among propositional connectives. It is in fact easy to see that the substitutional interpretation fails in contexts involving both quantifier and modal operators (or indeed any notions whose semantics involves a multiplicity of *possibilia*). An example from epistemic logic helps to illustrate what is involved here. We can use the meaning of the order of different operators as an *experimentum crucis*.

For the purpose, consider a sentence in which we "quantify in", for instance a sentence of the form

$$(\forall x)(\mathrm{A}(\mathrm{x}) \supset \mathrm{KB}(\mathrm{x})) \tag{7.1}$$

It might mean "each person in this town in known to be rich." This obviously does not imply "it is known that everybody in this town is rich", i.e.

$$\mathrm{K}\ (\forall x)(\mathrm{A}(\mathrm{x}) \supset \mathrm{B}(\mathrm{x})) \tag{7.2}$$

(Aristotle was already aware that the two are not equivalent; see *Analytica Priora* 67a16 ff. and *Analytica Posteriora* 74a30 ff.)

The reason for the non-equivalence is obvious: the denizens of this town of whom it is known that they are rich need not be known to exhaust its inhabitants. In other words, in some scenarios compatible with what is known, there exist other townsmen. And this is simply an instance of the dependence of the quantifier $(\forall x)$ on the virtual quantifier K. To neglect this dependence is to overlook the obvious difference between (7.1) and (7.2).

The kind of dependence of $(\forall x)$ on K that is involved here is as plain as the proverbial pikestaff (if you know what it is). The range of values of x in $(\forall x)$ is different in the different scenarios ("possible worlds") over which K ranges. The different knowledge worlds (scenarios compatible with what is known) need not have the same objects in them.

This affects the logical laws governing them. For instance, an inference from $(\forall x)$ KF[x] to K$(\forall x)$F[x] fails because there can in alternative knowledge worlds exist individuals that do not exist in the actual world, and an inference from P$(\exists$x)F[x] to $(\exists$x)PF[x] (where P expresses epistemic possibility) fails because possible individuals need not exist in the actual one. In all these cases, the sensitivity of meaning to the relative order of quantifiers and epistemic operators is a symptom of their dependence on each other.

Analogous remarks apply to modal operators in the narrow sense of the word, that is, possibility and necessity. There is no necessity (either logical or natural or nomic) that different worlds must have the same entities existing in them.

It should therefore be perfectly obvious that the substitutional interpretation of quantifiers fails in modal and epistemic contexts. Why then have several astute philosophers adopted it? The answer is that they have had some independent reason to do so. Characteristically, Bertrand Russell the acquaintance theorist and Ludwig Wittgenstein the author of the *Tractatus* believed that we should not assume the existence of any entities not given to us in direct experience. Hence such entities constitute the one and only range for our quantifiers. Alas, when Wittgenstein gave up (on October 11, 1929) the phenomenological languages that have objects of acquaintance as their universe of discourse, his reasons for clinging to the substitutional interpretation disappeared.

References

Frege, G. (1893–1903a). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. Jena: Herman Pohle.

Frege, G. (1893–1903b) *The basic laws of arithmetic* (M. Furth (Ed.) [partial translation of Frege 1893–1903], Trans.), Berkeley and Los Angeles: University of California Press.

Hintikka, J. (1996). *The principles of mathematics revisited*. Cambridge: Cambridge University Press.

Hintikka, J. (2004). On the different identities of identity: A historical and critical essay. In G. Fløistad (Ed.), *Language, meaning, interpretation* (pp. 117–139). Dordrecht: Kluwer Academic Publishers.

Hintikka, J. (forthcoming) If logic, definitions and the vicious circle principle.

Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16*, 83–94.

Author Biography

Jaakko Hintikka was Professor of Philosophy at Boston University. He was born in 1929 and educated in Finland. He defended his dissertation in 1953. In 1956–1959 Hintikka was Junior Fellow at Harvard. From 1959 until 2014 he held professorships at different institutions in Finland and in the US. Hintikka has created game-theoretical semantics and IF logic and made contributions to epistemic logic and epistemology, to inductive logic and the epistemology of induction, to the foundations of mathematics and to formal linguistics. He is also interested in the history of philosophy and has studied Aristotle, Descartes, Kant, Peirce and Wittgenstein among others. Hintikka has received six honorary doctorates. He has been honored by a volume in the series *Library of Living Philosophers* and by Grand Cross of the Order of the Lion of Finland. He died on the 12th of August 2015, when the volume was in print.

Frege and the Aristotelian Model of Science

Danielle Macbeth

Although profoundly influential for essentially the whole of philosophy's twenty-five hundred year history, the model of a science that is outlined in Aristotle's *Posterior Analytics* has recently been abandoned on the grounds that developments in mathematics and logic over the last 150 years have rendered it obsolete. Nor has anything appeared to take its place. As things stand, we have not even the outlines of an adequate understanding of the rationality of mathematics as a scientific practice. Frege, one of the last great defenders of the model *and* a key figure in the very developments that have been taken to spell its demise, can help us, I think, to begin to see a way forward. Frege, we will see, neither jettisons the model nor uncritically adopts it; instead he modifies it in important ways. So modified, I will suggest, the model can still serve as the basis for a cogent account of the rationality of mathematical practice.

According to the Aristotelian model, a system of concepts and judgments is a *science* just if it satisfies the following six desiderata:

- All those *concepts* and *judgments* concern a certain *domain of being*(*s*).
- Among the *concepts*, some are *primitive* and the rest are *defined* by appeal to those primitive concepts.
- Among the *judgments*, some are *primitive* and the remainder are *proven* as theorems from those primitive judgments.
- The *judgments* of the science are *true*, *necessary*, and *universal*.
- The *judgments* are *known to be true*, either directly or through proof.
- The *concepts* are *adequately known*, either directly or through definitions.[1]

Euclid's system as presented in the *Elements* has been throughout history the exemplar of a science in this sense, and so it is for Frege. He writes, for instance, in the introduction to *Grundgesetze* (1893, 2):

[1]This formulation largely follows that of de Jong and Betti (2010, 186).

D. Macbeth (✉)
Haverford College, Haverford, USA
e-mail: dmacbeth@haverford.edu

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_3

> The ideal of a strictly scientific method in mathematics, which I have here attempted to realize, and which might indeed by named after Euclid, I should like to describe as follows. It cannot be demanded that everything be proved, because that is impossible; but we can require that all propositions used without proof be expressly declared as such, so that we can see distinctly what the whole structure rests upon. After that we must try to diminish the number of these primitive laws as far as possible, by proving everything that can be proved. Furthermore, I demand—and in this I go beyond Euclid—that all methods of inference employed be specified in advance; otherwise we cannot be certain of satisfying the first requirement. This ideal I believe I have now essentially attained.[2]

But although Frege adopts this venerable conception of the scientific method, he also has a new agenda, one that greatly complicates his relationship to the model. That agenda is logicism.

One of the central aims of nineteenth century mathematics, evident in, for instance, Bolzano's work, was to show that intuition in the Kantian sense has no role to play in arithmetic and analysis.[3] And for Frege, as for many of his contemporaries, banishing intuition from arithmetic and analysis in this way seemed at the same time to suggest that arithmetic is after all a purely logical discipline, that its most basic principles are purely logical and all its concepts definable by appeal only to strictly logical primitives.[4] We read, for example, in Dedekind's famous 1887 essay "*Was sind und was sollen die Zahlen?*": "When I call arithmetic (algebra, analysis) only a part of logic, I state already that in my view the concept of number is totally independent of the representations or intuitions of space and time and that I take that concept as an immediate result of the laws of pure thought".[5] But Frege had two other reasons as well for thinking that logicism is true. The first concerns the domain of application of arithmetic, the fact that "the truths of arithmetic govern all that is numerable. This is the widest domain of all; for to it belongs not only the actual, not only the intuitable, but everything thinkable" (Frege 1884, §14; see also Frege 1885a, 112). If arithmetic, like logic, governs the widest possible domain then perhaps it really just is logic. Frege's second reason is the fact that, unlike the axioms of geometry, which can be denied without contradiction (however much that denial might conflict with our intuitions about the nature of space), it is not possible in the same way to deny a basic law of arithmetic: "try denying any one of them, and complete confusion ensues. Even to think at all seems no longer possible" (Frege 1884, §14). Again it seems natural to conclude that perhaps arithmetic is, after all, really logic. Frege's self-appointed task was definitively to show that arithmetic is "simply a development of logic, and every proposition of arithmetic a law of logic, albeit a derivative one" (Frege 1884, §87).

But logicism requires something more than showing that the fundamental axioms of arithmetic are purely logical. It requires showing as well that "there is no

---

[2]In "Logic in Mathematics" Frege (1914, 205) is less charitable, claiming that "Euclid had an inkling of this idea of a *system*; but he failed to realize it".

[3]See Coffa (1991, 27–9) for discussion.

[4]See Detlefson (2008, especially p. 187).

[5]Quoted in de Jong (1996, 302).

such thing as a peculiarly arithmetical mode of inference that cannot be reduced to the general inference-modes of logic", and that all concepts of arithmetic "be reducible to logic by means of definitions" (Frege 1885a, 113 and 114). Indeed, Frege (1885a, 114) suggests, it is only by such a reduction of the concepts of arithmetic to the concepts of logic that it is "possible to fulfill the first requirement of basing all modes of inference that appear to be peculiar to arithmetic on the general laws of logic".[6] In order, for example, to show that the law governing mathematical induction is a purely logical law, one needs to define the notion of following in a sequence by appeal only to strictly logical primitives. Frege does just that in the 1879 logic; he shows that the concept of following in a sequence—required in the reduction of "the argument from n to $(n + 1)$, which on the face of it is peculiar to mathematics, to the general laws of logic"—can be defined by appeal only to basic primitives of logic (Frege 1884, §80). The task was to complete such a reduction generally, for all axioms, all rules of inference, and all concepts that appear to be peculiar to the science of arithmetic, algebra, and analysis.

A commitment to logicism seems clearly to entail a commitment to something like the Aristotelian model of science insofar as the way to show that logicism is true is by systematically deriving all the basic truths of arithmetic from purely logical laws (by means of purely logical forms of inference), and defining all the basic concepts of arithmetic using only purely logical concepts. But one can be committed to the model without being committed to logicism. If logicism is false then arithmetic is a science with its own primitive notions and its own primitive laws. The way to show this, however, and thereby to settle the question of the fundamental nature of arithmetic, is, again, by settling the question of its most basic laws and therefore also its most basic concepts. Frege explains in "Logic in Mathematics" (1914, 204–5):

> Science demands that we prove whatever is susceptible of proof and that we do not rest until we come up against something unprovable. It must endeavour to make the circle of unprovable *primitive truths* as small as possible, for the whole of mathematics is contained in these primitive truths as in a kernel. Our only concern is to generate the whole of mathematics from this kernel. The essence of mathematics has to be defined by this kernel of truths, and until we have learnt what these primitive truths are, we cannot be clear about the nature of mathematics.

The mathematical demand for the utmost rigor, which includes the demand for proof where proof is possible, and the philosophical question of the true character of mathematical knowledge, whether purely logical or grounded in distinctively mathematical truths, result in one and the same requirement: "that the fundamental propositions of arithmetic should be proved, if in any way possible, with the utmost rigor; 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, and only when these are known shall we be able to answer our original questions" (Frege 1884, §4).

[6]See also Frege (1884, §4, 1914, 209).

Whatever the status of logicism, Frege clearly thinks that the demand for the utmost rigor, and hence proof where proof is possible, which requires in turn the analysis of concepts and the formulation of definitions of them, is a requirement of the science of mathematics: "in mathematics we must never rest content with the fact that something is obvious or that we are convinced of something, but we must strive to obtain a clear insight into the network of inferences that support our conviction" (Frege 1914, 205; see also 1884, §§90–91). But it can seem that the model inevitably commits us to something more, not only to rigor in mathematics but also to a false epistemology, in particular, to a demand for certainty or infallibility in mathematics. And this may well have been the way it was understood, for instance, by Descartes and by Kant. It is not, as Frege shows, constitutive of adherence to the model. One can, as Frege does, take the model to be the standard of scientific rationality in mathematics while being fallibilist about mathematical knowledge, that is, while thinking of mathematics as an "experimental" science, one whose foundations are not given and are never indubitably certain. Frege does claim, at the end of the introduction to *Grundgesetze*, that no one will be able either to produce a better system than his or to show that his principles lead to contradiction, but this is nothing more than an expression of his confidence that he has gotten things right. The test of one's logical convictions, Frege thinks, lies in what one can do with them, and it can always turn out (as it did for Frege) that unbeknownst to one, one has gotten something wrong. Russell's discovery of the flaw in Frege's Basic Law V, however personally devastating to Frege, nevertheless held out for him the promise of being the crucial first step in "a great advance in logic" (Frege 1902, 132). And it did so because, for Frege (1884, iii), "the first prerequisite for learning anything" is "the knowledge that we do not know".[7]

It is a familiar fact of intellectual life that one can think that one understands something clearly and distinctly even though in fact one does not. Because one can, the discovery of a problem that makes one *realize* that things are not clear, or distinct, can be of immense intellectual value, the necessary first step on the way to a better understanding. Such a discovery is possible, Frege thinks, in *all* domains of knowledge, including even logic. The laws of logic are truths (he thinks), and "what is true is true independently of our recognizing it as such. We can make mistakes" (Frege 1879b, 2). Of course, in most cases we have no reason to doubt the truth of this or that law of logic, the law of identity, say, that $a = a$. Such a law seems to us as manifestly true as anything could be. Because it does, we cannot imagine that it might be false. But, Frege suggests, we can imagine beings who doubt it, which is to say that we can imagine ourselves coming to have doubts about it, just as Russell brought Frege to have doubts about Basic Law V.[8] We cannot (now) imagine what

[7]See also Frege (1914, 221).

[8]In point of fact, Frege *had* had doubts but had managed to convince himself that the problems could be satisfactorily resolved. See my (2005, Chap. 5).

those doubts would be; if we could, we would have those doubts. But we can imagine having doubts. As Frege (1893, 15) puts the point: "the impossibility of our rejecting the law in question [the law of identity] hinders us not at all in supposing beings who do reject it; where it hinders us is in supposing that those beings are right in so doing, it hinders us in having doubts whether we or they are right". Because we have (now) no doubt at all that the law of identity is true, because we (now) can find no reason whatever to call it into question, we can have no doubt (now) that we are right to affirm it, and correlatively, that anyone who does not affirm it is wrong. But even so we can imagine beings who reject the law of identity, which, again, is just to say that we can imagine ourselves coming to reject it on grounds that are as yet unimaginable. Thus, for Frege as for Peirce, the first, and in a sense the only, law of reason is, in Peirce's (1992, 178) words, "that in order to learn you must desire to learn and in so desiring not to be satisfied with what you already incline to think". Again, mere conviction in mathematics is not enough. The ground of the conviction must be laid bare and it can be laid bare only by adhering to the demands of the model, by proving what can be proved and by defining all but a handful of primitive terms in terms of those primitive notions.

Frege furthermore provides us with an account of *how* it is that we can seem to be clear about something when in fact we are not, and as a constitutive element of that account, a conception of how it is that we can discover our error. Because, on Frege's view of language and cognition, all our awareness of what is objective is mediated by what he calls a sense, *Sinn*, which includes a mode of presentation of the thing in question, we can think we have grasped what is in cases in which the outlines of the sense are, in point of fact, confused and blurred. Indeed, Frege thinks (1884, vii), "often it is only after immense intellectual effort, which may have continued over centuries, that humanity at last succeeds in achieving knowledge of a concept in its pure form, in stripping off the irrelevant accretions which veil it from the eyes of the mind". We need, then, to distinguish between a concept and our knowledge of a concept, between "the logical and objective order" and "the psychological and historical order" (Frege 1885b, 136):

> a logical concept does not develop and it does not have a history … If we said instead 'history of attempts to grasp a concept' or 'history of the grasp of the concept', this would seem to me much more to the point; for a concept is something objective: we do not form it, nor does it form itself in us, but we seek to grasp it, and in the end we hope to have grasped it, though we may mistakenly have been looking for something when there was nothing. (Frege 1885b, 133.)

Concepts, at least mathematical concepts, are something objective that we can grasp and can fail to grasp. Furthermore, because, as we will later see in more detail, our grasp of a mathematical concept is through an inferentially articulated sense, we can discover by reasoning, by drawing inferences from the sense of a concept word, as least as far as we understand that sense, that we were after all mistaken about that sense. And this works because the contradictions that we fall into in the course of inquiry are

> created by treating as a concept something that was not a concept in the logical sense because it lacked a sharp boundary. In the search for a boundary line, the contradictions, as they emerged, brought to the attention of the searchers that the assumed boundary was still uncertain or blurred, or that it was not the one they had been searching for. So contradictions were indeed the driving force behind the search, but not contradictions in the concept; for these always carry with them a sharp boundary … The real driving force is the perception of a blurred boundary. (Frege 1885b, 134.)[9]

Getting clear on the sense of a concept word through such a process of "proof and refutation", inference and counter-inference, just is to come to grasp the concept in its pure form. The claims of infallibility and certainty that have historically attached to the model are in no way essential to it. If Frege is right, they should be jettisoned.

In its classical form, the model can seem to suggest that a constitutive feature of mathematical practice is the search for indubitable foundations, and so, it can seem, giving up the foundationalist enterprise in mathematics is tantamount to giving up the model.[10] Frege shows that this is just not so. One can be a thoroughgoing fallibilist as concerns our knowledge of the truths of mathematics (and logic), and nevertheless adhere to the model as the exemplar of scientific rationality. Indeed, it begins to seem that *really* to take seriously the fact that there is no certainty in mathematics *just is* to adhere to the model. But before we can pursue that thought, we need to consider yet another motivation for rejecting the model.

Axioms, on Frege's view, are basic principles that, assuming we have made no mistake, are and are known to be true. The received view of axioms in mathematics, derived from Hilbert, is that axioms are not truths but instead stipulations, implicit definitions that set various conditions on relational structures but are otherwise uninterpreted. In 1949, in his *Philosophy of Mathematics and Natural Science*, Weyl explained the view this way:

> An axiom system [is] a *logical mold of possible sciences* … One might have thought of calling an axiom system complete if in order to fix the meanings of the basic concepts present in them it is sufficient to require that the axioms be valid. But this ideal of uniqueness cannot be realized, for the result of an isomorphic mapping of a concrete interpretation is surely again a concrete interpretation … A science can determine its domain of investigation up to an isomorphic mapping. In particular, it remains quite indifferent as to the "essence" of its objects … The idea of isomorphism demarcates the self-evident boundary of cognition … Pure mathematics … develops the theory of logical "molds" without binding itself to one or the other among possible concrete interpretations … The axioms become *implicit definitions* of the basic concepts occurring in them.[11]

[9]The particular case Frege is referring to here concerns the notion of motion, but the point obviously generalizes.

[10]See, for example, Beth (1968, Chap. 2) and Rav (1999).

[11]Quoted in Shapiro (1997, 160).

On this view, the language of mathematics, within which its axioms are expressed, is to be conceived model-theoretically. This is not, we will see, the only possible conception one can have of that language.

Consider, for example, the familiar axioms of Hilbert's theory of order:

• $\sim(a < a)$	Irreflexivity
• $(a < b \ \& \ b < c) \supset a < c$	Transitivity
• $a < b \vee b < a \vee a = b$	Connectedness
• $(\forall x)(\exists y)(x < y) \ \& \ (\forall y)(\exists x)(x < y)$	No endpoints
• $(\forall x)(\forall y)(\exists z)(x < y \supset x < z < y)$	Denseness

Conceived model theoretically, these axioms are not true or false except relative to a model or interpretation. They implicitly define a certain formal structure and any collection of appropriately related objects can serve as a model for that structure. All such models are thus isomorphic. That is, as Hilbert explains in a letter to Frege (1899, 40–1):

> it is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep … and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things. In other words: any theory can always be applied to infinitely many systems of basic elements. One only needs to apply a reversible one-one transformation and lay it down that the axioms shall be correspondingly the same for the transformed things.

And if we do read the axioms of mathematics in this way, then the model will seem completely irrelevant to the practice of mathematics for the simple reason that mathematics so conceived is not really a science, an intellectual inquiry into objective mathematical truth, at all. It is instead a kind of a game answering to nothing outside itself. (And if mathematicians themselves recognize only a small subset of all the activities that might fit this description as good or interesting or significant mathematics, as they most certainly do, that can amount to nothing more than a subjective preference.)

Frege does not understand the language of mathematics model theoretically. Nor does he take it to be instead always already interpreted, as is sometimes assumed.[12] The uninterpreted/interpreted distinction, and with it the distinction of (logical) form and (semantic) content, simply has no application to mathematical language as Frege understands it.[13] As Frege understands it, generality is to be conceived not in

[12]See, for example, Goldfarb (1979), also van Heijenoort (1967).

[13]See my (2005) for an extended argument for this claim. In my (2012a), I argue that it is Frege's conception of mathematical language rather than that of the mathematical logician that we need if we are to understand the role of the practice of proving in coming to a better mathematical understanding. See also my (2014).

terms of the idea of quantifying over a domain of objects but instead by appeal to a second-level concept that is applied to a first-level concept. Thus for him Hilbert's axioms are in no way about any objects. Instead they concern the (first-level) relation that is designated by the sign '<', and they ascribe to that relation various fundamental properties, namely, irreflexivity, transitivity, and so on, that are designated in turn by collections of logical signs (the sign for generality, for the conditional, and so on) in the particular arrangements they have in the axioms.[14] The axioms, then, are to be conceived as articulating evident and basic truths about the relation designated by '<'. And once we have laid down such axioms for that basic relation, nothing about that relation beyond what is specified in the axioms can be appealed to in subsequent inferences. All that is presumed about that relation is what is expressed in the five axioms.

Notice now that it follows that any *other* first-level relation than that designated by '<' that also has the properties that are set out in the axioms will also have whatever properties can be derived from them. This is due to the nature of inference. If, for example, I can validly infer from the fact that, say, Felix is a cat that Felix is a mammal, then the inference remains valid if I substitute any other cat for Felix. The goodness of the inference depends in no way on the fact that it is about Felix. More generally, we can say that any actual inferences in particular cases are, similarly, instances of something more general. This is simply how it is with inference: as logicians have been emphasizing since Aristotle first founded the subject, inferences are valid in virtue of their form. And just the same is true in the case of Hilbert's axioms. If from the fact that a certain relation (namely, that designated by '<') has various properties (transitivity, and so on) it can be inferred that that relation also has certain other properties, then that inference remains valid if any other relation with those same properties is substituted for the original relation. What is spoken of as the peculiar abstractness or formal character of contemporary mathematical practice is thus understood by Frege instead in terms of concepts of higher-level properties and relations that are correctly ascribed to lower-level properties and relations.

According to the received view, the practice of mathematics was not merely radically transformed over the course of the nineteenth century; instead, it became something essentially *different*, something so different that the Aristotelian model that had hitherto served as the standard of rationality in the practice of mathematics could no longer be applied to it.[15] For Frege, the new practice, thought something essentially new, was still recognizably the same sort of investigation into mathematical truth that it had always been. The received view of the recent history of mathematics cannot, then, *explain* the Hilbertian conception of axioms. Instead it

[14]See Chaps. 3 and 4 of my (2005).

[15]See, for example, Azzouni's (2006, Chap. 6) discussion of what he takes to be the essentially differences between pre-twentieth century and twentieth (and twenty-first) century mathematics.

presupposes that conception. A more Fregean reading of that history would go roughly as follows.[16]

Mathematics began in ancient Greece with an understanding of its subject matter in terms of objects; what it was taken to be about was various sorts of arithmetical and geometrical objects, for instance, numbers conceived as collections of units, and various sorts of geometrical figures, circles, triangles, and so on. Then in the seventeenth century the practice of mathematics was fundamentally transformed, as was its subject matter. Instead of being concerned with objects, this new form of mathematical practice concerned itself instead with arithmetical relations, and in particular with functions that are expressible in the formula language of arithmetic and algebra, in equations.[17] In the nineteenth century mathematical practice again was transformed, and again its subject matter shifted. Now the concern was not with mathematical objects, nor even with mathematical functions. It was with mathematical *concepts*, concepts of objects such as numbers, concepts of properties of functions such as continuity, and concepts of structures such as that of a group or field.[18] The task for this new form of mathematical practice was to analyze and clarify all these various mathematical concepts, to define them explicitly (as best we can given our current understanding of what they entail), and to prove theorems deductively on the basis of those definitions. This was what Riemann, for example, was doing and it was this work that provided the context for Frege's work.[19] Frege explicitly and self-consciously developed his *Begriffsschrift*, his concept-script, as a formula language within which to do this new sort of mathematics, that is, to reason deductively from explicitly defined concepts. Although we today are used to reading Frege's language merely as a logic, as a notational variant of our own logical languages, *Begriffsschrift* was in fact to be a Leibnizian *characteristica* within which to *display the contents of concepts*, how they are built up out of primitive elements, and to do so in a way that enables rigorous deductive reasoning from those contents so displayed. As Frege already saw, no mere logic can serve such an expressive purpose.[20]

We have seen that if the model-theoretic account of mathematical language is correct then the Aristotelian model of mathematics as a science is not applicable to mathematics as currently practiced. We have, in that case, no account of the rationality of mathematics as a mode of intellectual inquiry. If, as Frege thinks, the model, suitably modified, does account for the rationality of mathematical practice, then the language cannot function model theoretically but must be understood instead as a Leibnizian *characteristica*. Indeed, we can say something even

[16]This history is explored in detail in my (2014).

[17]See my (2004), also Chap. 3 of my (2014).

[18]See my (2014, Chap. 5).

[19]See Laugwitz (1999), also Tappenden (2006).

[20]See my (2013), for an extended discussion of the expressive role of Frege's *Begriffsschrift*, and Chaps. 7 and 8 of my (2014) for an analysis of the reasoning from the contents of concepts that it enables.

stronger: if, as Frege thinks, mathematical practice is inherently fallible, an intrinsically open-ended process of self-correction, then it is only in terms of the model (or something very like the model) that we can understand the rationality of the enterprise of mathematics. What we need to understand, if only in outline here, is how it is that the model (or something very like it) ensures that we have the sort of cognitive control that is needed in order to make progress in mathematics conceived as a science.

According to Frege (Frege 1880/1, 34), errors in mathematics are almost invariably due to a lack of clarity about the relevant mathematical concepts: "almost all errors made in inference … have their roots in the imperfections of the concepts". And as we have already seen, we discover those imperfections by reasoning on the basis of concepts insofar as we understand them. But such a process of reasoning *can* reveal imperfections "only … if the content is not just indicated but is constructed out of its constituents by means of the same logical signs as are used in the computation. In that case, the computation must quickly bring to light any flaw in the concept formation" (Frege 1880/1, 35). This is precisely what Frege's concept-script enables, the expression of the *inferentially articulated* contents of concepts in a way that enables rigorous deductive proof and thereby a means to discover the flaws in our understanding of those concepts and to improve that understanding. In order to understand how this works, we need to consider more closely the structure of a science that can support such investigations.

We have seen already that according to Frege (1884, §2), "it is in the nature of mathematics always to prefer proof, where proof is possible" because "the aim of proof 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". A proof "serves to reveal logical relations between truths. That is why we already find in Euclid proofs of truths that appear to stand in no need of proof because they are obvious without one" (Frege 1914, 204). Indeed Frege thinks (1900, 157), that it is just this that "constitutes the value of mathematical knowledge": "not so much what is known as how it is known, not so much its subject-matter as the degree to which it is intellectually perspicuous and affords insight into its logical interrelations". But of course a proof must start somewhere; not everything can be proved. The task, then, is "to make the circle of unprovable *primitive truths* as small as possible", to discover some small collection of primitive truths from which all the others can be derived (Frege 1914, 204). And as Frege notes both in *Begriffsschrift*, §13, and in "Logic in Mathematics", often there is some leeway here insofar as two different axiomatizations of one domain of inquiry may be equally acceptable: "the possibility of one system does not necessarily rule out the possibility of an alternative system, so 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" (Frege 1914, 206).

Frege thinks of axioms as truths that are, or should be, immediately evident, that is, as truths about which we have no doubts (though, again, we can turn out to have been mistaken). As already noted, we today think of axioms instead as implicit definitions. It is worth emphasizing, again, that there is, even for Frege, something

right about this (although Frege would never so describe it), as we can see by reflecting a bit on Frege's practice in the 1879 logic. In Part I of that logic, Frege introduces all the primitive signs of his language, he explains his fundamental notions, and he sets out his one mode of inference. None of this, as he notes in the opening paragraph of Part II, can be expressed *in* his language because it forms the basis of all expression in the language. The elucidations of Part I belong, then, only to the antechamber of mathematics, the propaedeutics. They are necessary not to mathematics itself but in order to ensure that a reader has the same understanding of the basic notions of the system as its author.[21] It is in Part II that the development of the system properly begins. In Part II nine judgments are presented as the axioms that contain, as in a kernel, "all of the boundless number of laws that can be established" in logic (Frege 1879a, §13), and various theorems are derived from them using Frege's one mode of inference.

But what exactly is the relationship between the elucidations of Part I and the axioms and derived theorems of Part II? One aspect of that relationship is obvious. Because we need (on Frege's view) to recognize the truth of Frege's axioms, we need to know the thoughts they express, and in order to do that we need to grasp the senses of his primitive signs.[22] The elucidations of Part I are aimed at conveying the senses of Frege's primitive signs. Given, for example, what is expressed by the conditional stroke, and given what it is to judge, namely, to acknowledge the truth of a (true) thought, it follows that what is true is true on any condition you like (because what is true is true unconditionally). If, in other words, one has some acknowledged truth, some judgment, then one may infer that judgment on some condition, that is, add a condition to it, any condition you like. Frege's first axiom formulates this (valid) rule of inference in the form of a judgment. In the same way one needs to grasp the senses of Frege's other primitive signs in order to recognize the truth of his other axioms. But once one has done this, has come to see that the basic laws are evidently true, no further appeal to those senses may be made in proofs. For the purposes of proof, all that is granted as known regarding the primitive signs is what is made explicit in Frege's nine axioms. Because in Frege's concept-script everything necessary for a correct inference is fully expressed, nothing left to guessing (as he puts it in 1879, §3 of Part I), we can think of Frege's axioms as codifications of the fundamental inference potentials that are contained in his various signs. The axioms codify everything that is assumed to be known about the inferential significance of his signs. It is in this sense that they function rather like implicit definitions.

[21]See Frege's letter to Hilbert (1899, 36–7) and Frege (1914, 207).

[22]Frege did not introduce the *Sinn*/*Bedeutung* distinction until many years after the appearance of *Begriffsschrift*. Nevertheless, that distinction is already in play in the 1879 logic insofar as already in that logic Frege holds that we arrive at a concept word only by analyzing the thought expressed into function and argument. Even in *Begriffsschrift* sub-sentential expressions, whether simple or complex, designate something, have *Bedeutung* in addition to expressing a sense (*Sinn*), only in the context of a proposition and relative to a function/argument analysis.

Axioms, on Frege's view, are (or should be) truths that are immediately evident. Definitions are different insofar as they are not truths but instead stipulations: a definition "does not say, 'The right side of the equation has the same content as the left side.'; but, 'They are to have the same content'" (Frege 1879a, §24). What a definition stipulates is that some newly introduced, hitherto meaningless sign has precisely the same meaning (*Bedeutung*) as some collection of primitive, and perhaps also already defined, signs. Frege claims that the two also express the same sense (*Sinn*), that the newly introduced sign is merely an abbreviation for the complex sign used in its definition, but in fact this cannot be right given that he also thinks that proofs of theorems from (fruitful) definitions can constitute real extensions of our knowledge. We will come back to this. For now we simply assume that for Frege definitions stipulate sameness of meaning (*Bedeutung*) but not also sameness in the sense (*Sinn*) expressed. Once the stipulation has been made, an identity judgment immediately follows, albeit one that is, in light of the stipulation, utterly trivial, not only immediately evident (*einleuchtend*), as an axiom is, but self-evident (*selbstverständlich*).

Although definitions are not judgments, judgments, albeit utterly trivial ones, do follow from definitions, and those judgments, Frege thinks, can serve as the starting points of proofs. Thus, although in Part II of *Begriffsschrift*, Frege proves various theorems on the basis of axioms given his one rule of inference, in Part III of that work the theorems that are proven are proven *not* on the basis of those axioms and theorems together with his definitions but *only* from his definitions. Relative to those definitions as they figure in the derivations in Part III, the axioms and theorems of Part II function not as premises from which to reason but instead as rules according to which to reason.[23] In effect, what in Part II appear as *judgments*, as premises and conclusions of inferences, function in Part III instead as *inference licenses*. There are two fundamentally different sorts of cases. First, there are rules governing one-premise inferences, where a one-premise inference is an inference that serves merely to transform some judgment (derived ultimately from a definition) in some way, say, by reordering the conditions in it. But Frege recognizes also two-premise inferences, inferences that serve to *combine* content from two judgments (ultimately derived from definitions). These sorts of inferences are governed by the rule of hypothetical syllogism, or some variant of it.[24] It is illuminating to compare Frege's practice here with that of eighteenth century algebraic problem solving.

Basic algebra can be codified in a collection of familiar basic laws such as the commutative and associative laws of addition and multiplication, the distributive law, and so on. On the basis of these laws other formulae can be derived such as the familiar law that $(a + b)^2 = a^2 + 2ab + b^2$. These derived laws can then be used just

[23]As Frege says in the long Boole essay, the proof of theorem 133 is "*from* the definitions of the concepts following in a series, and of many-oneness *by means of* my primitive laws" (1880/1, 38; emphasis added).

[24]Frege regularly distinguishes between one- and two-premise inferences, for example, in his (1879, §6), and in (1914, 204). A much more extensive discussion of these two sorts of inference, which I call linear and joining inferences, can be found in my (2012b, and 2014, §7.3).

as the basic laws are used, to license transformations of equations by putting identicals for identicals. So far, then, what we have is comparable to the basic and derived laws of Part II of the 1879 logic. But there is in algebra also a general rule that equals can be put for equals, and it is this rule (analogous to the rule of hypothetical syllogism in Frege's system) that enables one to prove interesting results such as, for example, Euler's theorem that $e^{ix} = \cos(x) + i \sin(x)$. The basic and derived laws of algebra enable one to rewrite formulae in various ways so that eventually, using the rule that equals can be put for equals, one can combine what at first seemed wholly unrelated, namely, in our example of Euler's theorem, the exponential function and two trigonometric functions. Similarly, in Frege's system, most of the basic and derived rules serve to enable one to rewrite formulae derived from definitions in various ways so that eventually, using some variety of hypothetical syllogism (also derived in Part II), one can combine contents that are at first wholly unrelated, derived from two different definitions.[25] The principal theorem that Frege proves in Part III of *Begriffsschrift*, theorem 133, is not as interesting as Euler's theorem; nevertheless, it does establish on purely logical grounds a theorem that would otherwise seem to depend on some intuition about sequences. Unlike the theorems that are derived from axioms in Part II Frege's proof of theorem 133 constitutes a real extension of our knowledge, just as Frege claims in *Grundlagen* (1884, §91): "From this proof it can be seen that propositions which extend our knowledge can have analytic judgements for their content."[26]

In the Aristotelian model as traditionally understood, definitions seem to have nothing whatever to do with the proof of theorems. There are the primitive and defined terms, on the one hand, and the primitive and derived judgments, on the other, and no indication of any sort of relationship between the two. This is unsurprising given that in Euclid's system definitions have no role to play in demonstrations. Definitions in Euclid, like elucidations in Frege, belong to the preamble or antechamber, not to the actual system of mathematics.[27] What formulates the contents of the concepts of concern to Euclid for the purposes of demonstration in Euclid are not definitions but instead drawn diagrams.[28] Mathematics as it has been practiced since the nineteenth century is in this regard essentially different from Euclid's practice insofar as proofs in contemporary mathematical practice *do* take explicitly formulated definitions as their starting points. (See, for example, any textbook of modern, abstract algebra.) And Frege's system, we have seen, reflects this. Part III of *Begriffsschrift*—the aim of which, Frege tells us, is to "give a general idea" of how to conduct proofs in Frege's concept-script (1879a, §23)—introduces four definitions and proves on the basis of those definitions a series of theorems

[25]In my (2011, and 2014, §7.3) I discuss this example in more detail and further develop the parallels between reasoning in Euler and reasoning in Frege.

[26]I develop and defend this claim that the derivation of theorem 133 is at once deductive and ampliative, a real extension of our knowledge, in a preliminary way in my (2012b) and in more detail in Chap. 8 of my (2014).

[27]See Netz (1999, §2.2).

[28]See my (2010), also Chap. 2 of my (2014).

culminating in theorem 133 showing a logical relation among three of Frege's four defined concepts. As Frege shows by example, definitions can enable the discovery of logical relations among (defined) concepts by means of proof.

In Frege's practice, the theorems that are derived from axioms of logic are very different from theorems that are derived from definitions. But the two sorts of derivations are not merely independent of one another, as primitive and defined concepts appear to be independent of primitive and derived theorems in the classical model. In Frege's system, axioms and the theorems that are derived from them provide inference licenses for proofs that take definitions as their starting points. And this is possible because precisely the same signs that appear in the axioms and theorems are used also in the formulation of the definitions: "the content … is constructed out of its constituents by means of the same logical signs as are used in the computation" (Frege 1880/1, 35). Now in Frege's *Begriffsschrift* example, even the "constituents" are strictly logical; *all* of the primitive notions in the system of the 1879 logic belong to logic. But even were the language to be enriched with the addition of primitive signs of arithmetic, the same would be true. The new primitives would require the addition of new axioms codifying the fundamental inferential significance of those primitives. But they would also enable the formulation of definitions of various mathematical concepts, concepts such as that of continuity, of prime number, or of being a common multiple, all of which, and more, are formulated in Frege's concept-script together with some signs from arithmetic in Frege's early essay "Boole's logical Calculus and the Concept-script". Proofs of (significant) theorems in the new system would begin with such definitions and proceed in accordance with the rules codified in the axioms of the system together with the theorems that are derived from those axioms. There would, then, remain a fundamental division between, on the one hand, axioms and theorems derived from them, all of which function in the system overall as rules governing inference, and on the other, definitions and theorems derived from those definitions. Only the latter sorts of theorems, theorems proven on the basis of definitions, would constitute real extensions of our knowledge.

Neither proofs without definitions, that is, proofs merely from axioms, nor, more obviously, definitions without proofs, can extend our knowledge according to Frege. Only definitions and proofs *together*, that is, proofs *of* theorems *from* definitions, can reveal something new. And, as I have argued in detail elsewhere, they can reveal something new precisely because definitions provide both simple unanalyzable signs for the concepts of interest and also an articulation, in the definiens, of their inference potential, their senses. The simple signs ensure that the derived theorem is about the concepts of interest, that it is not merely a logical consequence of the axioms; and the definitions of those concepts, which lay out their significance for inference, make possible the proof involving those concepts, the proof of a theorem showing that the defined concepts first appearing in different definitions have, demonstrably, a certain logical relationship one to another.[29] Were

[29]See my (2012b) and Chaps. 7 and 8 of my (2014).

(simple) defined signs merely abbreviations of the complex signs used to define them, this would be unintelligible; one would not in that case be able to distinguish in any principled way between theorems derived from axioms and theorems derived from definitions; all theorems would be nothing more than logical consequences of the axioms. That a theorem derived on the basis of definitions is essentially different from a theorem derived from axioms is intelligible only if we take the signs flanking the identity sign in a definition to have (by stipulation) the same *Bedeutung* but also to differ in the senses expressed. The definiens is a complex sign that mirrors, or maps, the logical articulation of the sense; and because it is complex, rules of inference can be applied to it in the context of a proposition. The defined sign, the definiendum, is simple; it has no logical articulation and no rules of inference can be applied to it. The two signs, definiens and definiendum, thus have radically different inferential significances. They express different senses. It follows directly that definitions do not merely introduce abbreviations.

According to the model as traditionally understood a science comprises concepts, both primitive and defined, and judgments, both primitive and derived. No indication is given of the relationship, if any, between concepts and judgments. Frege's practice, which (though less explicitly) is essentially that of mathematicians beginning in the nineteenth century and continuing still today, exhibits a much more complex structure. In Frege's practice, there are, first, axioms setting out the fundamental inferential significances of the primitive signs, as well as various theorems that follow from those axioms by a recognized rule of inference. Then there are the definitions of concepts, which are stipulations that introduce some new simple sign and assign it a meaning that is given in the definiens by way of an inferentially articulated sense. These definitions are, furthermore, formulated using just the same signs as are employed in the formulation of the axioms. And finally there are the theorems that follow from the definitions by means of the axioms and theorems derived from those axioms. In these latter derivations, the definitions (more exactly the judgments that derive directly from them) serve as premises; the axioms, and theorems derived from those axioms, serve instead as rules of inference governing the passage from the definitions to, ultimately, the theorem one aims to prove. The resultant proof (of a theorem on the basis of definitions) is completely rigorous and gap-free. And because it is, it is manifest on what the theorem depends and by what means it is justified. Thus, as Frege remarks in the introduction to *Grundgesetze* (1893, 3), "if anyone should find anything defective, he must be able to state precisely where, according to him, the error lies: in the Basic Laws, in the Definitions, in the Rules, or in the application of the rules at a definite point. If we find everything in order, then we have accurate knowledge of the grounds upon which each individual theorem is based".

Because everything is made explicit in reasoning in Frege's system, both the starting points and the rules governing inferences, errors are more easily detected and where no errors are found, though this is no *guarantee* that problems will not later come to light, one has very reasonable assurance that the theorem is true and good reason to think one has discovered the nature of its grounds, whether in logic alone or in the basic laws of some special science. It is in just this that the rationality

of the endeavor consists. The inquiry is, in Sellars' (1956, §38) words, "rational, not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once". Because it makes everything maximally explicit and hence available to critically reflective scrutiny and criticism, the method gives one *cognitive control* over the domain of inquiry. Although it does not guarantee that mistakes will not be made, does not guarantee that what seems to be clear and distinct really is clear and distinct, it ensures as far as possible that such errors as there are will, sooner or later, come to light and be corrected. The practice is both fruitful and maximally robust. It is, as mathematics is and has always been, self-consciously rational.

On the Aristotelian model as traditionally understood, knowledge of the basic concepts and judgments of a science is somehow immediate and certain. Frege, we have seen, rejects this aspect of the model. For him, mathematics is an "experimental" science whose foundations are not given but instead must be discovered by a process of proof and refutation, that is, self-correction. Axiomatizations and explicit definitions are, on this understanding, not so much the end product of science as a means, a vehicle of discovery. If a contradiction is derived then one knows that something is wrong, either with one's axioms or with one's definitions, that there is something one has not adequately understood, that one needs to make some corrections. But we have seen that Frege modifies the classical model in another way as well. Where the model has concepts, primitive and defined, and judgments, primitive and derived, Frege has instead (1) axioms and theorems of logic, (2) definitions formulated using the same signs as are used in the axioms and theorems of logic, and (3) theorems derived from definitions. In Frege's modified model, theorems derived on the basis of (fruitful) definitions are essentially different from theorems derived from axioms alone and must be distinguished from them. This, I have suggested, is something radically new, something that is characteristic in particular of mathematical practice as it emerged over the course of the nineteenth century and continues today. Although they are not as rigorous in their practice as Frege is and they do not use a formula language of the sort Frege developed (although they could), what mathematicians do is to define concepts and derive theorems based on their definitions, showing thereby various logical relations that obtain among the concepts so defined. What Frege helps us to see is in what the rationality of this practice consists. As he shows, the Aristotelian model of science, updated to reflect developments within mathematics, can still today provide a viable and compelling image of scientific rationality by showing, if only in the broadest outline, how it is that we achieve, and maintain, cognitive control in our mathematical investigations.[30]

[30]An earlier, shorter version of this essay was presented at the first international meeting of the Association for the Philosophy of Mathematical Practice (APMP) in Brussels, Belgium, December 2010. I am grateful to the participants for very useful discussion.

References

Azzouni, J. (2006). *Tracking reason: Proof, consequence, and truth*. Oxford: Oxford University Press.

Beth, E. W. (1968). *The foundations of mathematics: A study in the philosophy of science*. Amsterdam: North Holland.

Coffa, J. A. (1991). *The semantic tradition from Kant to Carnap*. L. Wessels (Ed.). Cambridge: Cambridge University Press.

de Jong, W. R. (1996). Gottlob Frege and the analytic-synthetic distinction within the framework of the Aristotelian model of science. *Kant-Studien, 87*, 290–324.

de Jong, W. R., & Betti, Arianna. (2010). The classical model of science: A millennia-old model of scientific rationality. *Synthese, 174*, 185–203.

Detlefson, M. (2008). Purity as an ideal of proof. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 179–197). Oxford: Oxford University Press.

Frege, G. (1879a). *Conceptual notation, a formula language of pure thought modelled upon the formula language of arithmetic*. In T. W. Bynum (Ed. and Trans., 1972), *Conceptual notation and related articles* (pp. 101–203). Oxford: Clarendon Press.

Frege, G. (1879b). Logic. In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Posthumous writings* (P. Long & R. White, Trans., 1979), pp. 1–8. Chicago: University of Chicago Press.

Frege, G. (1880/1). Boole's logical calculus and the concept-script. In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Posthumous writings* (P. Long & R. White, Trans., 1979) pp. 9–46. Chicago: University of Chicago Press.

Frege, G. (1884). *Foundations of arithmetic* (J. L. Austin, Trans., 1980). Evanston, IL.: Northwestern University Press.

Frege, G. (1885a). On formal theories of arithmetic. In B. McGuinness (Ed.), *Collected papers on mathematics, logic, and philosophy* (M. Black et al., Trans., 1984), pp. 112–121. Oxford: Basil Blackwell.

Frege, G. (1885b). On the law of inertia. In B. McGuinness (Ed.), *Collected papers on mathematics, logic, and philosophy* (M. Black et al., Trans., 1984), pp. 123–136. Oxford: Basil Blackwell.

Frege, G. (1893). *The basic laws of arithmetic: Exposition of the system* M. Firth (Ed. and Trans., 1964, with an introduction). Berkeley and Los Angeles: University of California Press.

Frege, G. (1899). Letter to David Hilbert, 27 December. In G. Gabriel et al. (Ed.), *Philosophical and mathematical correspondence* (H. Kaal, Trans., 1980), pp. 34–38. Chicago: Chicago University Press.

Frege, G. (1900). Logical defects in mathematics. In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Posthumous writings* (P. Long & R. White, Trans., 1979), pp. 157–166. Chicago: University of Chicago Press.

Frege, G. (1902). Letter to Bertrand Russell, 22 June. In G. Gabriel et al. (Eds.) *Philosophical and mathematical correspondence* (H. Kaal, Trans.), pp. 131–133, 1980. Chicago: Chicago University Press.

Frege, G. (1914). Logic in mathematics. In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Posthumous writings* (P. Long & R. White, Trans.), pp. 203–250, 1979. Chicago: University of Chicago Press.

Goldfarb, W. (1979). Logic in the twenties: The nature of the quantifier. *Journal of Symbolic Logic, 44*, 351–368.

Hilbert, D. (1899). Letter to Frege, 29 December. In G. Gabriel et al. (Eds.), G. Frege, *Philosophical and mathematical correspondence* (H. Kaal Trans.), pp. 38–41, 1980, Chicago: Chicago University Press).

Laugwitz, D. (1999). *Bernhard Riemann 1826–1866: Turning points in the conception of mathematics* (Abe Shenitzer, Trans.). Boston, Basel, and Berlin: Birkhäuser.

Macbeth, D. (2004). Viète, Descartes, and the emergence of modern mathematics. *Graduate Faculty Philosophy Journal, 25*, 87–117.

Macbeth, D. (2005). *Frege's logic*. Cambridge, MA: Harvard University Press.
Macbeth, D. (2010). Diagrammatic reasoning in Euclid's *Elements*. In B. Van Kerkhove, J. De Vuyst, & J. P. Van Bendegem (Eds.), *Philosophical perspectives on mathematical practice, texts in philosophy* (Vol. 12, pp. 235–267). London: College Publications.
Macbeth, D. (2011). Seeing how it goes: Paper-and pencil reasoning in mathematical practice. *Philosophia Mathematica, 20*(1), 58–85.
Macbeth, D. (2012a). Proof and understanding in mathematical practice. *Philosophia Scientiae, 16* (1), 29–54.
Macbeth, D. (2012b). Diagrammatic reasoning in Frege's *Begriffsschrift*. *Synthese, 186*, 289–314.
Macbeth, D. (2013). Writing reason. *Logique et Analyse, 221*, 25–44.
Macbeth, D. (2014). *Realizing reason: A narrative of truth and knowing*. Oxford: Oxford University Press.
Netz, R. (1999). *The shaping of deduction in Greek mathematics*. Cambridge: Cambridge University Press.
Peirce, C. S. (1992). *Reasoning and the logic of things: The Cambridge conference lectures of 1898*. In K. L. Ketner (Ed.), Cambridge, MA: Harvard University Press.
Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica, 7*, 5–41.
Sellars, W. (1956). Empiricism and the philosophy of mind. In *Science, perception, and reality* (pp. 127–196). London: Routledge and Kegan Paul (1963).
Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. Oxford: Oxford University Press.
Tappenden, J. (2006). The Riemannian background to Frege's philosophy. In J. Ferreirós & J. J. Gray (Eds.), *The architecture of modern mathematics: Essays in history and philosophy* (pp. 97–132). Oxford: Oxford University Press.
Van Heijenoort, J. (1967). Logic as calculus and logic as language. *Synthese* 17:324–330.

Author Biography

Danielle Macbeth is T. Wistar Brown Professor of Philosophy at Haverford College in Pennsylvania and the author of *Frege's Logic* (Harvard University Press, 2005) and of *Realizing Reason: A Narrative of Truth and Knowing* (Oxford University Press, 2014). She has also published on a variety of issues in the history and philosophy of mathematics, philosophy of language, philosophy of mind, pragmatism, and other topics. Macbeth was a Fellow at the Center for Advanced Study in the Behavioral Sciences in 2002–3, and has been the recipient of both an American Council of Learned Societies (ACLS) Burkhardt Fellowship and a Fellowship from the National Endowment for the Humanities (NEH).

On the Nature, Status, and Proof of Hume's Principle in Frege's Logicist Project

Matthias Schirn

Introduction: Hume's Principle, Basic Law V and Cardinal Arithmetic

Thanks to the pioneering work on Frege's foundations of arithmetic by Wright (1983) and Boolos (1987), Hume's Principle has attracted much attention in the philosophy of mathematics during approximately the last twenty five years. In the project of neo-logicism or neo-Fregeanism, launched by Wright (1983), further developed by himself and Bob Hale (see Hale and Wright 2001) and widely discussed until today, Hume's Principle has even gained the status of an icon or at least it may seem so. In *Die Grundlagen der Arithmetik* (*The Foundations of Arithmetic*), §§62–3, Frege introduces Hume's Principle as a second-order abstraction principle and attempts to define the cardinality operator "the number which belongs to the concept φ" ("$N_x\varphi(x)$") in terms of it: $N_xF(x) = N_xG(x) \leftrightarrow Eq_x(F(x), G(x))$, in words: the number that belongs to the concept F is equal to the number that belongs to the concept G if and only if F and G are equinumerous, that is, if and only if the objects falling under F can be correlated one-to-one with those falling under G. The attempt failed by Frege's own lights.

In *Grundlagen*, and also in some of his subsequent work, Frege criticizes several methods of abstraction in philosophy and in logic, but refrains from characterizing the transition from right to left in Hume's Principle or that in his preferred example from geometry "The direction of line a is identical with the direction of line b if and only if a and b are parallel" by using the term "abstraction". Analogous remarks apply to a comment that he makes on Basic Law V in the second volume of *Grundgesetze der Arithmetik* (*Basic Laws of Arithmetic*), §146 in the course of discussing the issue of whether his introduction of value-ranges via Axiom V can

M. Schirn (✉)
Ludwig Maximilians University Munich, Munich, Germany
e-mail: matthias.schirn@lrz.uni-muenchen.de

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_4

be called a creation. Frege probably thought that due to his rejection of what he considered to be misguided methods of abstraction the term "abstraction" had acquired a negative meaning, at least for himself (cf. Frege 1884, §21, §34, §45, §§49–51, 1967, pp. 164 f., 186 ff., 214 ff., 240–261, 324 ff.). If so, then in the light of his logicist manifesto he would possibly have deemed the term inappropriate for referring to what we nowadays call *Fregean abstraction* and *abstraction principles*.[1] In fact, in the course of criticizing Cantor's method of obtaining the ordinal or cardinal number of a set via a single or a double act of abstraction, he says that the verb "to abstract" is a psychological expression and, as such, to be avoided in mathematics. Thus, with the exception of *we* call Fregean abstraction, it seems that Frege regarded abstraction in its manifold guises as a thorn in the flesh.

In *Grundlagen*, §64, Frege explains the move from right to left in the example from geometry, that is, the step of abstraction, in terms of an act of splitting up or dissecting a content in a way different from the orginal way, which is supposed to give us a new concept. In my opinion, this characterization is an infelicitous choice of phrasing. Closer examination reveals that it is not only doubtful in itself but also has nothing to do with Fregean abstraction correctly understood. Perhaps Frege became aware of this inadequacy later in his work, as his succinct description of the act of abstraction in Axiom V might indicate (cf. Frege 1903, §146). So, it is after all not surprising that Frege's metaphorical way of describing abstraction in *Grundlagen*, §64 has provoked misunderstanding in the literature.

At first glance, there is nothing spectacular about Hume's Principle, especially when we look at it in isolation, independently of a mathematical theory it which it may figure as a definition or as a provable theorem or as an axiom. However, as soon as we focus our attention on the details of Frege's attempt to demonstrate that cardinal arithmetic is a branch of logic—informally and only sketchily in *Grundlagen* and formally and definitively in *Grundgesetze*—we see that Hume's Principle is the pivot of the proofs of the basic laws of number theory. In *Grundgesetze*, Basic Law V—$(\dot{\varepsilon}f(\varepsilon) = \dot{\alpha}g(\alpha)) \leftrightarrow (\forall x(f(x) \leftrightarrow g(x)))$, in words: the value-range of f is identical with the value-range of g if and only if f and g are coextensive—which is the exact structural of Hume's Principle, is *formally* only needed for framing the explicit definition of the cardinality operator in purely logical terms and for establishing the logical status of Hume's Principle by deriving it from that definition. However, this is not to deny that Frege regarded Basic Law V as the linchpin of his logicist enterprise. No doubt, he did regard it as such

[1] In his letter to Russell of 28.7.1902, Frege makes some comments on abstraction principles and takes the transformation of the equivalence relation of geometrical similarity into an identity of shapes of, say, triangles as an example. He mentions that this is what Russell perhaps calls "définition par abstraction". In Chapter XI "Definition of cardinal numbers" of his *Principles of Mathematics* (Russell 1903), Russell does apply the term "definition by abstraction", but makes it clear that such a definition "suffers from an absolutely fatal formal defect: it does not show that only one object satisfies the definition. Thus instead of obtaining *one* common property of similar classes, which is *the* number of the classes in question, we obtain a *class* of such properties, with no means of deciding how many terms this class contains" (p. 114).

precisely because he had chosen it as a means of introducing logical objects of the required fundamental and irreducible kind and of providing a means that enables us to have a non-intuitive cognitive access to them.[2] Yet he probably knew that Axiom V could provide the appropriate epistemic contact with value-ranges of functions only if he was able to resolve a serious problem arising from his semantic stipulation in §3. This stipulation failed to fix completely the references of canonical value-range terms.

As to the primitive objects of logic, the True and the False, Frege felt entitled to assume that everybody, including the sceptic about truth and falsity, is already familiar with them in his or her ordinary practice of judging and asserting and that he could therefore safely rely on them in his logic. In particular, unlike the value-ranges, the truth-values did not give rise to any indeterminacy or underdetermination. Since in Frege's view all numbers had to be identified with value-ranges (cf. Frege 1893, §9) in order to establish arithmetic—at least number theory and real analysis—as a branch of logic, one could say that he construed value-ranges as the target objects par excellence of his reductionist enterprise, while the truth-values were placed in the ground floor of first-order and second-order logic together with functions, concepts and relations.[3] In any event, it is indisputable that Hume's Principle is the fulcrum of the proofs of the basic laws of cardinal arithmetic and as such plays a fundamental role in Frege's overall foundational project. Plainly, the roles that in *Grundgesetze* he assigns to Axiom V and Hume's Principle regarding the foundation of number theory differ essentially from one and another, but it is not incorrect to say that the axiom and the principle were designed to work hand in hand.

Guided by the title of this collection of articles by different hands and also considering the limitation of space, I shall place my study almost entirely in a Fregean context, that is, I shall mainly ignore the discussion of the status and role of Hume's Principle, for example, in a neo-Fregean context. The only notable exception is Section "The Nature of Abstraction: A Critical Assessment of *Grundlagen*, §64" in which I criticize not only Frege, but also Wright and his fellow combatants for neo-logicism. In recent years, neo-Fregeanism or neo-logicism has more or less become a branch of research in its own right. In my view, it is in fact largely independent of Frege's own logicist concerns.

Section "The Julius Caesar Problem in *Grundlagen*—A Brief Characterization" of this essay is likewise peparatory for the ensuing discussion of specific issues

[2]Frege's reaction to Russell's discovery of a contradiction in his letter to Russell of 22.6.1902 speaks volumes with respect to the key role that Axiom V was intended to play in his logicist project. "I must give some further thought to the matter. It is all the more serious as the collapse of my Law V seems to undermine not only the foundations of my arithmetic, but the only possible foundation of arithmetic as such" (Frege 1976, p. 213).

[3]In Frege (1893, §10), Frege also identifies the True and the False with special value-ranges, but his motive for doing this is entirely different from that for the (projected) identification of numbers of any kind with value-ranges. In fact, the identification of the truth-values with their unit classes is intended to remove, in a first essential step, the referential indeterminacy of value-range names deriving from the semantic stipulation made in §3 and later to be enshrined in the formal version of Basic Law V.

concerning Hume's Principle. In it, I briefly characterize the Julius Caesar problem that Frege faces in *Grundlagen* when he explores the possibility of defining the cardinality operator contextually via Hume's Principle. In Section "Analyticity", I consider the options that Frege might have had to establish the analyticity of Hume's Principle, bearing in mind that with its analytic or non-analytic status his envisaged purely logical foundation of cardinal arithmetic stands or falls. Section "Thought Identity and Hume's Principle" is devoted to the two criteria of thought identity that Frege states in 1906 and to their application to Hume's Principle. In Section "The Nature of Abstraction: A Critical Assessment of *Grundlagen*, §64", I shall subject to scrutiny his characterization of abstraction in *Grundlagen*, §64 as well as the currently widespread use of the terms "recarving" and "reconceptualization" with respect to Fregean abstraction. Section "Frege's Proof of Hume's Principle" is devoted to the formal details of Frege's proof of Hume's Principle. I begin by considering his proof sketch in *Grundlagen* and subsequently reconstruct in modern notation essential parts of the formal proof in *Grundgesetz*e with special emphasis on the analysis which precedes each construction in the proof. In Section "Equinumerosity and Coextensiveness: Hume's Principle and Basic Law V Again", I focus my attention on the criteria of identity embodied in Hume's Principle and in Axiom V, equinumerosity and coextensiveness. In Section "Julius Caesar and Cardinal Numbers—A Brief Comparison Between *Grundlagen* and *Grundgesetze*", I comment on the Caesar problem arising from Hume's Principle in *Grundlagen* and analyze the reasons for its absence in this form in *Grundgesetze.* I conclude with reflections on the introduction of the cardinals and the reals by abstraction in the context of Frege's logicism.

The Julius Caesar Problem in *Grundlagen*—A Brief Characterization

In *Grundlagen*, §65, Frege tentatively introduces the direction operator (and by analogy the cardinality operator) by means of the following definition: the sentence "line a is parallel to line b" *is to mean the same as* (*sei gleichbedeutend mit*) "the direction of line a is identical with the direction of line b" (the sentence "the concepts F and G are equinumerous" *is to mean the same as* (*sei gleichbedeutend mit*) "the cardinal number that belongs to the concept F is identical with the cardinal number that belongs to the concept G"). In Schirn (2014b), I have argued, by invoking especially *Grundlagen,* §104, that in *Grundlagen* a contextual definition that presents itself in the guise of an abstraction principle stipulates that one side is to mean the same as the other in the sense that both sides shall have the same judgeable content, which, in Frege's later terminology, amounts to stipulating that they shall express the same thought.

In *Grundlagen*, §66, Frege rejects the proposed contextual definition of the cardinality operator "$N_x\varphi(x)$" in terms of Hume's Principle on the grounds that this

definition gives rise to the Julius Caesar problem (cf. *Grundlagen*, §66).[4] The criterion of identity for cardinal numbers inherent in Hume's Principle, namely the equinumerosity of F and G, does not provide a means that enables us to decide whether, say, the number of planets is identical with Julius Caesar, or speaking more generally: the tentative contextual definition of "$N_x\varphi(x)$" does not place us in a position to determine the truth-value of an equation of the form "$N_xF(x) = t$" and, hence, does not fix uniquely the reference of "$N_x\varphi(x)$" or of a numerical term "$N_xF(x)$". Here, "t" is an arbitrary singular term (for example, "the smallest prime number", "the successor of 1 in the natural series of numbers" or "2") differing in form from "$N_xG(x)$". Understood in this way, the Caesar objection is undoubtedly a semantic problem. Note that Frege does not yet use the term "semantic". He speaks of a third *logical* doubt to which the tentative contextual definition of "$N_x\varphi(x)$" gives rise. In any event, the Caesar problem concerning cardinal numbers in §66 affects Frege's logicist project as outlined in *Grundlagen* profoundly.[5] By contrast, its precursor in §56 patently misses its mark and therefore has no decisive impact on his project of introducing cardinal numbers as logical objects (cf. Schirn 2003, 2014b). In §56, Frege raises a Caesar objection in the course of analyzing the alleged shortcomings of his proposed inductive definition of the natural numbers. So much for the Julius Caesar problem at this stage. I shall return to it repeatedly in subsequent sections, especially in Section "Julius Caesar and Cardinal Numbers—A Brief Comparison Between *Grundlagen* and *Grundgesetze*". In what follows, I shall discuss Frege's notion of analyticity with an eye to Hume's Principle.

Analyticity[6]

At the outset of the Preface to his *Begriffsschrift* of 1879, Frege stresses that the firmest method of proof is the purely logical one, which appeals only to the laws on which all knowledge rests. I take it that the laws in question are the primitive truths of logic and the purely logical rules of inference. Frege divides all truths which require a proof for their justification into two kinds. The proof of a truth of the first kind can proceed in a purely logical fashion, while the proof of a truth of the second kind must be supported by empirical facts. In a moment, when I turn to

[4]Recall that Frege defines the relation of equinumerosity in second-order logic in terms of one-to-one correlation.

[5]If Frege were to carry out his plan in *Grundlagen*, §104 to define, in a first step and after the fashion of the attempted contextual definition of the cardinality operator, fractions, irrational and complex numbers by using second-order or higher-order abstraction, he would face a whole family of Caesar problems each of which is supposed to be resolved by framing an appropriate explicit definition for these numbers in terms of extensions of concepts.

[6]This section is a slightly revised and enlarged version of Section "The Nature of Abstraction: A Critical Assessment of *Grundlagen*, §64" in M. Schirn, 'Frege's Logicism and the Neo-Fregean Project', Axiomathes 24 (2014), 207–243.

Grundlagen, §3, we shall see that Frege's early division or classification of truths that stand in need of proof amounts to his later distinction between analytic and (synthetic-) a posteriori truths in §3. It is striking that in the Preface he identifies the grounding or justification of a truth without much ado with its proof. Admittedly, gapless proofs count as a safe, straightforwardly verifiable kind of grounding sentences, but they are not the only possible mode of justifying our acknowledgement of the truth of a sentence or a thought. Frege was aware of this as several of his other writings show. One of his seventeen key sentences on logic reads as follows (Frege 1969, p. 190, cf. p. 183):

> We justify a judgement either by going back to truths that have been recognized already or without having recourse to other judgements. Only the first case, inference, is the concern of logic.

According to Frege, deductive inference is to judge by being aware of other truths as grounds of justification. In his fragment 'Logik' (I), he emphasizes that deductive inference cannot be the only mode of justifying truths (Frege 1969, p. 3):

> Now the grounds which justify the recognition of a truth often reside in other truths which have already been acknowledged. Yet if there are any truths acknowledged 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 require justification at all.

In 'Logik' (I), Frege assigns the task of investigating non-deductive forms of justification to epistemology. Thus logic and epistemology are here put on a par insofar as both disciplines are concerned with justifying grounds of judgements. Unfortunately, Frege does not mention any other forms of grounding or justifying truths besides deductive proof. In particular, he does not say that epistemology can provide a non-deductive justification of a primitive law of logic. It therefore remains unclear on what the justification of truths (if there are any), which are capable and (or) in need of justification, but resist justification through deductive proof, is supposed to rest. Yet it seems that if there were no such truths, epistemology, as characterized by Frege, would lack a proper domain of investigation. For he can hardly see its task in furnishing justifying grounds for truths which do not stand in need of justification. Note that in 'Logik' (I) Frege does not explicitly claim or require the existence of truths that need neither deductive nor non-deductive justification. He seems to suggest though that there must be certain truths whose recognition does not depend on a proof. From his viewpoint, every deductive proof in a theory T must proceed from premises whose truth has already been acknowledged so that the proof convinces us of the truth of the proved thought, the conclusion. However, the first premises of inferences in T must be truths which do not stand in need of proof and must be explicitly characterized as such in order to avoid any circularity in the conduct of proof.

Let us now turn to *Grundlagen*. In §3, Frege defines the notion of analyticity in terms of the notion of deductive proof. This is quite in the spirit of the opening remarks in the Preface to *Begriffsschrift*, bearing in mind that in *Begriffsschrift* he

does not yet use the notion of analyticity on his own account.[7] According to *Grundlagen*, §3, a truth is analytic if (and only if) it can be proved entirely from general logical laws (that neither need nor are capable of proof) and definitions. It is hereby presupposed that we also take into account all propositions on which the admissibility of a definition rests. Regrettably, Frege does not mention any such proposition. I think that what he has in mind here, though perhaps only allusively, are principles for putting forward correct definitions. It seems that in *Grundlagen* Frege did not yet rely on a theory of definition with clear-cut principles. He states such principles only as late as in Frege (1893, §33) (cf. Frege 1903, §§ 56–67). In any event, in *Grundlagen* his practice of framing definitions was not yet guided by his later principle of the simplicity of the *definiendum* and his prohibition on piecemeal definitions. His treatment of the tentative contextual definition of the direction operator (or of the cardinality operator) in *Grundlagen* provides ample evidence for this.

In *Grundlagen*, Frege defines the concept of analyticity only for truths that can be proved. No provision is made for the first premises of the deductive proof (of an arithmetical truth), namely the primitive laws of logic figuring as axioms in a theory T and the definitions framed in T. The primitive truths of logic qua axioms are unprovable only relative to T, as Frege certainly knew; see, for example, Frege (1969, pp. 221 f.). Even a primitive truth of logic may be provable in a theory T^*, but due to the self-evidence it is supposed to possess it does not require proof. While in *Begriffsschrift* $a = a$ is laid down as an axiom, in Frege (1893) it appears as a theorem. In Frege (1893, §50), Frege comments on "$a = a$" qua theorem of his logical calculus as follows: "Although this sentence is by our explanation of the equality-sign obvious [*selbstverständlich*], it is nonetheless worth seeing how it can be developed out of (III)." (III) is Basic Law III: $g(a = b) \rightarrow g(\forall \mathfrak{f}(\mathfrak{f}(b) \rightarrow \mathfrak{f}(a)))$, in words: the truth-value $g(\forall \mathfrak{f}(\mathfrak{f}(b) \rightarrow \mathfrak{f}(a)))$ falls under every concept under which the truth-value $a = b$ falls. "$a = a$" does not need proof, because it is obvious. Deriving "$a = a$" nonetheless from Basic Law III is not pointless, although the proof does not amount to furnishing a deductive justification for "$a = a$". This accords well with Frege's remark that it is worth the effort to show how "$a = a$" can be inferred from Basic Law III despite the obviousness of this sentence. The proof

[7] In *Begriffsschrift*, Frege employs the terms "synthetic" and "analytic" only in two places, firstly in the course of discussing the notion of identity of content ("*Inhaltsgleichheit*") in §8, and secondly when he comes to explain the nature and purpose of definitions framed in his concept-script in §24. Frege takes as an example the definition of a hereditary property in a series. He refers to it as formula (69). Since in formula (69) we do not acknowledge a judgeable content as true, but make a stipulation, (69) is "not a judgement; and consequently, to use a Kantian expression, also *not a synthetic judgement*" (Frege 1879, p. 56). Once the content of the *definiens* has been bestowed upon the *definiendum* the definition is immediately turned into an analytic judgement; for "we can only get out what was put into the new symbols in the first place" (p. 56). Thus, Frege uses here the term "analytic" along Kantian lines, namely as an equivalent for "epistemically trivial judgement". It is true that according to Frege's definition of "analytic truth" in *Grundlagen*, §3 the definitions turned into assertoric sentences are likewise to be regarded as analytic. Yet in contrast to Kant, he argues that there are analytic truths containing valuable extensions of our knowledge.

even of an obvious truth in a theory *T* may serve to gain a deeper insight into the inferential links that exist between the truths of *T* and, hence, into the logical structure of *T*. In 'Logik in der Mathematik' ('Logic in Mathematics') (Frege 1969, p. 220), Frege argues in this vein:

> A proof does not only serve to convince us of the truth of what is proved; it also serves to reveal logical relations between truths. This is why Euclid already proved truths that appear to need no proof, because they are evident without one.

Now, I assume that Frege, had his attention been drawn to the omission in his definition of the concept of analyticity, would have characterized both the primitive laws or axioms of logic and the definitions as analytic. In Frege (1893) and in his later work on the foundations of arithmetic, the term "analytic" has disappeared from his active vocabulary, but he nowhere explains why he has dispensed with it. However, it seems likely that in his view the terms "analytic truth" and "logical truth" are coextensional.[8] The opposite assumption does not carry an awful lot of conviction.[9]

As I pointed out in Section "Introduction: Hume's Principle, Basic Law V and Cardinal Arithmetic", Frege considered Hume's Principle to be the linchpin of the proofs of the basic laws of cardinal arithmetic. In the light of his logicist credo, it was therefore mandatory to establish the analytic or logical nature of this fundamental number-theoretic principle. Considering also the need to define the cardinality operator in purely logical terms after having explored an initial, unsuccessful definition of the finite cardinals as objects (Frege 1884, §55), it could seem that in *Grundlagen* Frege had basically three options to accomplish this. (1) To use Hume's Principle as a contextual definition of the cardinality function, bearing in mind that equinumerosity is conceptually prior to the cardinality function and, hence, must be defined in the first place in purely logical terms; (2) to derive Hume's Principle from an explicit definition of the cardinality operator, which was designed to conform to this principle and whose *definiens* was likewise couched in purely logical terms; (3) to lay down Hume's Principle as a logical axiom governing the cardinality operator as a primitive term of the concept-script later to be employed in the formal execution of the logicist programme. Note that after 1890 option (1) was no longer viable for Frege, quite independently of the impact that the Caesar problem had on the acceptability of the tentative contextual definition of the

[8] In Schirn (2016b), I analyze in detail Frege's use of "analytic" in the period 1879–1892.

[9] See the controversy between Boolos (1997) and Wright (1999) concerning the analyticity or non-analyticity of Hume's Principle. In a sense, one might say that their different views result from the distinct notions of analyticity which underlie their arguments; see Schirn (2006). See in this connection also Ebert (2008) on the dispute between the neo-Fregean (represented by Wright) and the "epistemic rejectionist" (represented by Boolos). Boolos jettisons the key idea of neo-Fregeanism, namely that Hume's Principle and certain other abstraction principles are analytic truths, despite the fact that they involve specific ontological commitments. Ebert calls Boolos's position *epistemic* rejectionism, since Boolos refuses to accept Wright's idea that Hume's Principle *can be known* on analytic grounds, or as Wright has also put it: that our *knowledge* of number theory may be regarded as deriving a priori from Hume's Principle.

cardinality operator in terms of Hume's Principle. The theory of definition in Frege (1893), which was based on clear-cut principles and tailored to the concept-script, did not permit any definition of this kind. According to this theory, contextual definitions offend against the principle of the simplicity of the *definiendum* and possibly against the principle of completeness as well. The latter prohibits piecemeal definitions.[10]

Frege dismisses option (1), due to the impact of the Caesar problem and the apparently hopeless venture of removing it by making an additional stipulation which was consistent with the contextual definition of the cardinality operator and did not rest on intuition or experience. Establishing the analyticity of Hume's Principle by deriving it from an explicit definition of the cardinal number that belongs to the concept *F* in terms of an equivalence class of precisely that equivalence relation which Hume's Principle displays as the criterion of identity for cardinal numbers, is what Frege actually does both in *Grundlagen* and in *Grundgesetze*. It seems that during that time this was, from his point of view, by far the most promising option. Firstly, unlike option (3), it did not require that Hume's Principle be self-evident.[11] Yet the self-evidence of any Fregean abstraction principle qua equation of the form "$a = b$" (or qua equivalence statement "$a \leftrightarrow b$")—where "a" itself is an equation of the form "$a = b$" and "b" is a sentence or statement expressing that a certain equivalence relation holds between the members of a given domain—implies that "a" and "b" have the same judgeable content (express the same thought, respectively). And from the fact that such an equation can be transformed into an equation of the form "$a = a$" without any change of content or sense it follows that "$a = b$" does not contain real knowledge.[12] Admittedly, it is true that in his attempt to define the cardinality operator contextually Frege *stipulates* that the two sides of Hume's Principle shall be "*gleichbedeutend*". Recall that in my view "shall be *gleichbedeutend*" is most likely intended to be tantamount to "shall have the same judgeable content", and not to "referential". But we should not infer from this that Frege continued regarding the two sides as "*gleichbedeutend*" in this sense once he came to realize that the initial role

[10]See Frege's argument for rejecting contextual definitions in Frege (1903, §66).

[11]It is clear that for Frege the self-evidence of a truth cannot serve as a general criterion of analyticity. On the one hand, he grants that there are non-evident sentences which are analytic truths, such as, for example, the equation "125,664 + 37,863 = 163,527", provided that the logicist programme has been successfully carried out for cardinal arithmetic. On the other hand, Frege acknowledges the existence of self-evident, but non-analytic truths, such as the axioms of Euclidean geometry. For a true statement "$a = b$" to be analytic in Frege's sense, the identity of the sense(s) of "a" and "b" is a sufficient condition, but it is not a necessary one.

[12]For a detailed discussion of this issue see Schirn (2016a, Sects. 5 and 6). To avoid misunderstanding here, let me mention that in *Grundlagen* Frege did not yet construe declarative sentences as truth-value names, as names of the True or the False, and hence as a special kind of (complex) proper names that can appropriately flank "=" on both sides. Yet if Hume's Principle is considered to be a statement of the form "$a \leftrightarrow b$", the argument above for its epistemic triviality, if "a" and "b" have the same judgeable content or express the same thought, would equally apply.

of Hume's Principle qua definition of the cardinality operator had to be abandoned and that this principle had to be established as a provable theorem.

Secondly, adopting option (2) enabled Frege to provide a deductive justification for Hume's Principle, to secure in this way its requisite analytic or purely logical status and thus to stick to the aim of providing a logical foundation of cardinal arithmetic. There is of course a striking difference between the *Grundlagen* and the *Grundgesetze* cases. As was remarked earlier, in *Grundlagen* Frege simply assumes that one knows what the extension of a concept is and, furthermore, takes the logical nature of extensions of (second-level) concepts tacitly for granted. It is this assumption which casts a gloom over his logicist project as outlined in *Grundlagen*. Clearly, when Frege was writing *Grundlagen* he could not rely on a commonly accepted view of the nature of extensions of concepts, let alone of the nature of numbers. By contrast, in *Grundgesetze*, §3 he introduces extensions of first-level concepts (more generally: value-ranges of monadic first-level functions) via a metalinguistic stipulation later to be incorporated in the formal concept-script version of Axiom V (cf. §20), and, thus, from his point of view, as logical objects in a sound manner. In saying this, I disregard the inconsistency of Basic Law V of which Frege was unaware in Frege (1893). I further ignore the re-emergence of the Caesar problem from his semantic stipulation in §3. However, Frege probably thought that he had succeeded in endowing canonical value-range terms with a unique reference by determining for every primitive first-level function when introducing it which values it receives for value-ranges as arguments, just as for all other fitting arguments.[13]

Thought Identity and Hume's Principle[14]

In *Grundlagen*, when Frege comes to define the cardinality operator tentatively in terms of Hume's Principle, he stipulates that the two sides of Hume's Principle shall have the same judgeable content (or thought). In this section, I try to figure out whether in the light of the two criteria of thought identity, which Frege formulated in 1906, the two sides of Hume's Principle should in fact be considered to express the same thought. If that were the case, then the analytic nature of Hume's Principle would be secured in the context of Frege's theory of sense and reference. But the consequence involved in this would be unpalatable for Frege: the stigma of epistemic triviality.

Frege formulates his first criterion of thought identity in a letter to Husserl. He proceeds from the assumption that neither of the members of a given pair of

[13]For details see Schirn (2016c).

[14]This section is a slightly revised and enlarged version of Sect. 6 in M. Schirn, 'Frege's Logicism and the Neo-Fregean Project', Axiomathes 24 (2014), 207–243.

sentences contains a logically evident sense component. The criterion, henceforth referred to as "CRIT 1", is this (Frege 1976, p. 105 f.):

> If both the assumption that the content of [a sentence] *A* is false and that of *B* true, and the assumption that the content of *A* is true and that of *B* false lead to a logical contradiction, and if this can be established without knowing whether the content of *A* or *B* is true or false, and without requiring other than purely logical laws for this purpose, then nothing can belong to the content of *A*, insofar as it is capable of being judged true or false, which does not also belong to the content of *B* … Equally, under our assumption, nothing can belong to the content of *B*, insofar as it is capable of being judged true or false, which does not also belong to the content of *A*.

In a posthumously published piece entitled 'Kurze Übersicht meiner logischen Lehren' ('A Brief Survey of my Logical Doctrines') [1906] (Frege 1969, p. 213), Frege states a different criterion, henceforth referred to as "CRIT 2":

> Two sentences *A* and *B* can stand in such a relation that anyone who acknowledges the content of *A* as true must without further ado acknowledge that of *B* as true and, conversely, that anyone who acknowledges the content of *B* as true, must also immediately acknowledge that of *A* (*equipollence*), where it is presupposed that there is no difficulty in grasping the contents of *A* and *B*.

CRIT 1 is formulated in logical terms, CRIT 2 in epistemic. CRIT 1 embodies a notion of sense or thought which is akin to the notion of conceptual content of *Begriffsschrift*. In this book, Frege states indirectly a criterion of identity for the conceptual contents of sentences or judgements (henceforth abbreviated as "CICC"). According to CICC, two sentences S_1 und S_2 have the same conceptual content, if the consequences that can be derived from S_1 in combination with certain other sentences $T_1,\ldots,T_n$ can always be derived also from S_2 in combination with $T_1,\ldots,T_n$. Frege mentions "At Plataea the Greeks defeated the Persians" and "At Plataea the Persians were defeated by the Greeks" as an example of two sentences that have the same conceptual content, and adds that in his concept-script one need not distinguish between any two sentences S_1 und S_2 that satisfy CICC. As regards CRIT 2, we see that it incorporates a notion of sense or thought which is of a finer texture than the one contained in CRIT 1. When we apply CRIT 1 to Hume's Principle, it seems natural to conclude that its two sides express the same thought.[15] However, the matter proves to be difficult, when we apply CRIT 2 to Hume's Principle.

CRIT 2 it is not free of vagueness. Firstly, it is not quite clear what Frege means by the presupposition that there is no difficulty in grasping the content of *A* and that of *B,* because he spares himself the trouble of spelling this out. Under what conditions could we attribute to a person P a full grasp of the thought(s) expressed by the two sides of Hume's Principle? Surely, P must know that asserting the

[15]See also Dummett's comment on CRIT 1 in Dummett (1981, p. 324). It remains unclear why the view involved in this criterion, namely that two analytically equivalent sentences express the same thought, should be irreconcilable with Frege's ideas about sense as stated in other writings. Dummett owes us a plausible explanation why there should be a serious conflict.

equinumerosity of two (non-empty) concepts F and G amounts to the assertion that the objects falling under F can be correlated one-to-one to the objects falling under G. Furthermore, P must have a grasp of one-to-one correlation. Yet does a full grasp of Hume's Principle also require that P knows at least some of the consequences that can be drawn from it, say, when it is embedded in standard axiomatic second-order logic? And how about the ontological commitment that Hume's Principle entails? Must P also have a clear idea about this?

Secondly, it is likewise not clear how we should understand the modal term "must" in this context. Applied to Hume's Principle, it seems to me that what Frege has in mind is that anyone who acknowledges the thought expressed by the right-hand side of Hume's Principle as true, but refuses to acknowledge the thought expressed by its left-hand side as true, and conversely, faces a logical contradiction. It is, however, perfectly conceivable that someone who is aware of the fact that the transition from right to left in Hume's Principle involves a possibly dangerous ontological assumption, is reluctant to acknowledge the left-hand side immediately as true, although he did not hesitate to acknowledge its right-hand side as true (for a certain pair of concepts F and G). While the right-hand side concerns only the one-to-one correlation between first-level concepts F and G and makes no demand on the domain of objects over which the first-order variables range, the left-hand side, by contrast, requires that the first-order domain comprises at least denumerably many objects, assuming that we look at Hume's Principle from the point of view of the platonist.[16] And if someone passes from the acknowledgement of the truth of the (cognitive) content of A to the acknowledgement of the truth of the (cognitive) content of B only on (possibly much) reflection, CRIT 2 would not be satisfied. Moreover, it is conceivable that someone recognizes "standard" instances of Hume's Principle as true, but hesitates or refuses to acknowledge special instances such as "$N_x(x = x) = N_xFCN(x) \leftrightarrow Eq_x(x = x, FCN(x))$" as true. (I use here "$FCN$" as an abbreviation for "finite cardinal number".) As Boolos (1987, p. 16) shows, this equation is an undecidable sentence in the formal system FA (Frege arithmetic), which is axiomatic second-order logic plus Hume's Principle. It is true in some models of FA, but false in others. There are further examples of equations of the form "$N_xF(x) = N_xG(x)$" whose truth-value is undecidable (for us); see in this respect Heck (1999, p. 262). The upshot is that the two sides of Hume's Principle might come out as expressing the same thought according to CRIT 1, but they might also be considered to express different thoughts according to CRIT 2.

We do not know whether Frege was aware of the tension or even mismatch that appears to exist between CRIT 1 and CRIT 2. He presumably regarded the two criteria as equivalent, because it seems that he formulated them in the same year. In any case, it is hard to fathom why he should have deemed it necessary to work with two distinct notions of thought in his logic. If we account for several other remarks

[16]Ebert (2008) argues in detail that Hume's Principle involves the existence of infinitely many objects only when it is combined with additional metaphysical assumptions concerning the existence of properties. He claims that in an Aristotelian universe, where there are no empty concepts, Hume's Principle will never inflate an originally finite domain to an infinite one.

that Frege makes on the difference of sentence sense, the suspicion grows stronger that he lacked perhaps a coherent view of thought identity and thought difference. Here is an example that may indicate this lack of coherence. According to CRIT 1, the thought expressed by "$2^2 = 4$" would be the same as that expressed by "$2 + 2 = 4$", and this would possibly also hold according to CRIT 2. Yet Frege explicitly contends that the two equations express different thoughts (see Frege 1893, p. 7, 1976, p. 235).

The Nature of Abstraction: A Critical Assessment of *Grundlagen*, §64[17]

In this section, I comment on Frege's description of abstraction in *Grundlagen*, §64 and the use of the terms "recarving" and "reconceptualization" in the relevant literature on Fregean abstraction and neo-logicism.[18] The matter is not of marginal interest, neither with respect to the aim of clarifying Frege's characterization of abstraction in *Grundlagen*, nor with regard to Wright's neo-Fregean approach to the foundations of cardinal arithmetic. If we take Wright's statement that an instance of the left-hand side of Hume's Principle is meant to embody a *recarving* or *reconceptualization* of the type of state of affairs depicted on the right (cf. Wright 1997, 1999) at face value, and if in using these terms he has indeed different decompositions of the same content or proposition in mind, then it could seem that the neo-logicist project is bound to fail from the very outset. The reason is that neither first-order nor second- or higher-order abstraction principles involve in any plausible sense a recarving or different decompositions of one and the same content.

In Wright (1999, p. 209), Wright explains: "Numbers are, rather, like directions, the output of a distinctive kind of reconceptualization of an epistemologically prior species of truth". Expressing it in this rather vague fashion, Wright seems to rely on Frege's way of describing first-order abstraction in *Grundlagen*, §64. So, let us first look at Frege's description.

> The judgement "The straight line a is parallel to the straight line b", in symbols $a \parallel b$, can be construed as an equation. If we do this, we obtain the concept of direction, and say: the direction of the straight line a is identical with the direction of the straight line b. We replace the symbol $\parallel$ with the more general symbol $=$, by distributing the particular content of the former symbol to a and to b. We split up the content in a way different from the original way, and thereby obtain a new concept.

At first glance, it could seem that what Frege has in mind here regarding the move from right (a) to left (b) in an abstraction principle are different dissections or decompositions—an original and a new one—of one and the same judgeable

[17]This section is a slightly revised and marginally enlarged version of Sect. 9 in M. Schirn, 'Frege's Logicism and the Neo-Fregean Project', Axiomathes 24 (2014), 207–243.

[18]In *Grundlagen*, §64, Frege uses the word "*zerspalten*".

content. The new decomposition is said to yield a new concept, namely the concept of a direction. Seen in this way, the process described in the passage from §64 may be reminiscent of Frege's thesis, stated by him already in his early writings, that "we start out from judgements and their contents, and not from concepts. We only allow the formation of concepts to proceed from judgements [via decomposition]" (Frege 1969, pp. 17 f.; cf. Frege 1976, p. 164). The thesis is frequently called the thesis of the priority of judgeable contents over concepts. In his 'Aufzeichnungen für Ludwig Darmstaedter' ('Notes for Ludwig Darmstaedter') of 1919 Frege states the priority thesis as follows: "I do not begin with concepts and put them together to form a thought or judgement; rather, I arrive at the parts of a thought by decomposing it" (Frege 1969, p. 273). Dummett even goes so far as to claim in one place (in Dummett 1991, p. 173) that decomposition of a content is surely the model for Frege's contention in §64 that it is by splitting up the content of (a) or of (b) in a new way that we attain the concept of a direction or that of a cardinal number. However, appearances can be deceptive; on closer examination, Dummett's claim seems to be unjustified. If it were true, we should think that Frege was confused when he wrote §64. It is in *Grundlagen*, §70, not in §64, that he describes the true method of forming (first-level) concepts and relations through analysis of a judgeable content.

In order to gain a clearer idea of what is involved in Fregean abstraction as characterized in *Grundlagen*, §64 and what is not, I begin by stating succinctly the rules that in *Grundgesetze*, §26 Frege lays down for the construction of well-formed expressions (proper names and function-names) from the primitive function-names of his concept-script, which are well-formed by stipulation. I do this mainly because in §26 he formulates in a precise manner and in appropriate terminology what in *Begriffsschrift*, 'Booles rechnende Logik und die Begriffsschrift' ('Boole's Logical Calculus and the Concept-Script') and in *Grundlagen* he had still expressed in terms of judgement and judgeable content and concept, without clearly specifying decomposition or analysis as an operation that is applied in the first place to sentences (or more generally: to complex expressions), and therefore is a purely syntactic device. The central question regarding *Grundlagen*, §64 that we must raise and answer is this: Can Frege's talk of splitting up the content of the right-hand side of an abstraction principle in a way different from the original way be understood along the lines of his method of extracting function-names by means of gap formation to be explained below?

The formation rules for the concept-script expressions are these: (1) the rule of insertion which licences the formation of (a) complex proper names (function-value names) by inserting a fitting argument-expression into the argument-place of a monadic function-name of first or of second or of third level and (b) complex monadic function-names of first level by inserting a proper name into the ζ- or ξ-argument-place of a dyadic first-level function-name; (2) the rules of gap formation (as I call them) which govern the construction of (i) monadic first-level function-names, (ii) dyadic first-level function-names and (iii) monadic second-level function-names (with an argument-place of the second or of the third kind, as the case may be) by removing some or all occurrences of a proper name

either from a more complex proper name (case (i) in the standard version) or from a first-level monadic function-name (case (ii)) or by removing some or all occurrences of a monadic first-level function-name from a proper name (case (iii)) and by marking the resulting gap(s) as an argument-place of the appropriate kind (cf. Frege 1893, §26).[19] The operation of the extraction of function-names by means of gap formation is supposed to go hand in hand with the process of analyzing a complex expression into parts (call this *analysis*) and the parallel process of splitting up or decomposing the complex sense of the expression into simpler sense components (call this *decomposition*).[20]

Take the right-hand side (a) "$Eq_x(F(x), G(x))$" of Hume's Principle and recall Frege's characterization of abstraction in *Grundlagen*, §64. We can analyze (a) into the sign for a second-level equivalence relation "$Eq_x(\varphi(x), \psi(x))$" and the two first-level predicates "*F*" and "*G*", and thanks to this decompose the thought expressed by (a) into the doubly unsaturated sense of "$Eq_x(\varphi(x), \psi(x))$" and the senses of "*F*" and "*G*". Let us call this the *standard analysis* of (a) and the *standard decomposition* of the thought expressed by (a). The standard analysis of (a) proceeds as follows: In a first step, we apply Frege's third gap formation rule to (a), which in his terminology after 1891 is a proper name or a name of a truth-value.[21] We thus obtain the name of a second-level concept, namely "$Eq_x(\varphi(x), G(x))$". From this we can remove, in a second step, the occurrence of "*G*" and obtain "$Eq_x(\varphi(x), \psi(x))$". Note that in Frege (1893, §26), Frege does not state a fourth gap formation rule which licenses the formation of a dyadic second-level function-name by removing the occurrence(s) of one-place first-level function-name from a monadic second-level function name (of which the former forms a part) and by marking the resulting gap as an argument-place of the second kind. But I think that he could have added such a rule without further ado, had he considered it to be essential for the syntax of his concept-script. To be sure, in the case of Frege's paradigm of a first-order abstraction principle "$D(a) = D(b) \leftrightarrow a \parallel b$" we could apply his first gap formation rule to "$a \parallel b$" and then his second rule to "$\xi \parallel b$" (or to "a $\parallel \zeta$") and in this way analyze "$a \parallel b$" into "*a*", "*b*" and "$\xi \parallel \zeta$". I presume that Frege has this analysis in mind when in §64 he speaks of the *original* way of splitting up the content which "$a \parallel b$" and "$D(a) = D(b)$" are said to share.

[19]Argument-places of the first kind are suitable for the insertion of proper names; argument-places of the second kind are suitable for the insertion of monadic first-level function-names; and argument-places of the third kind are suitable for the insertion of dyadic first-level function-names; cf. Frege (1893, §23).

[20]Not to be confused with a similar distinction drawn by Dummett in Dummett (1981, pp. 333 f.).

[21]In Schirn (2016c, d), I introduce the term "truth-value name", which Frege does not use; as far as I can see, he only uses the term "name of a truth-value". By a truth-value name I understand a function-value name that has the syntactic structure of a declarative sentence, and hence does not only refer to one of the two truth-values, but also expresses a thought. By contrast, the value-range name "$\grave{\varepsilon}(\varepsilon = (\varepsilon = \varepsilon))$" or the definite description "$\backslash\grave{\varepsilon}$ (—ε)" do not express a thought, although, due to Frege's stipulations in *Grundgesetze*, both names refer to the True. They are names of a truth-value, but they are not truth-value names qua sentences. Both names express a complex, non-propositional sense.

Let us turn again to Hume's Principle and ask: What, in the light of §64, is the new decomposition of the content of (a) "$Eq_x(F(x), G(x))$" supposed to be? Following Frege's stipulation that (a) and (b) "$N_xF(x) = N_xG(x)$" of Hume's Principle shall be *gleichbedeutend*, that is, shall have the same judgeable content according to my interpretation, we can decompose this content or thought into the senses of the two singular terms "$N_xF(x)$" and "$N_xG(x)$" and the sense of the two-place predicate "$\xi = \zeta$". This decomposition rests on the standard analysis of (b). The latter proceeds by applying the first gap formation rule to "$N_xF(x) = N_xG(x)$" and then by applying the second gap formation rule to "$N_xF(x) = \zeta$" . If we apply Frege's third gap formation role to "$N_xF(x) = N_xG(x)$", we obtain with "$N_x\varphi(x) = N_xG(x)$" (or "$N_xF(x) = N_x\varphi(x)$") the name of a monadic second-level function with an argument-place of the second kind and in this way analyze (b) and, moreover, decompose the thought expressed by (b) in a non-standard fashion. Of course, "$N_xG(x)$" (or "$N_xF(x)$" could be further analyzed into "G" and "$N_x\varphi(x)$" (or into "F" and "$N_x\varphi(x)$") by applying again Frege's third gap formation rule.

To close this section, I want to make three further points..

(i) It is evident that neither the act of analyzing (a) and (b) in the standard fashion via gap formation, nor the standard decompositions of the judgeable content or thought that (a) and (b) are said to share by stipulation, have anything in common with Fregean abstraction understood in the proper sense (cf. Frege 1903, §§146f.). If in *Grundlagen* Frege thought that they have, he must have been utterly confused when he was writing §64. As soon as the transition from (a) to (b) has been made by way of acknowledging that F and G have something intrinsically in common, namely their cardinal number, or, in other words, by assigning the same object (cardinal number) to the equinumerous concepts F and G, the act of abstraction has already been carried out. Plainly, any subsequent standard or non-standard analysis of (b) and, by the same token, any subsequent standard or non-standard decomposition of the thought expressed by (b) is irrelevant regarding the step of abstraction in Hume's Principle. Note that both the standard and non-standard decompositions of the thought expressed by (b) are inseparably interwoven with the standard and non-standard analysis of (b). As Frege puts it figuratively in several places (Frege 1967, p. 378, 1969, pp. 243, 262, 275, 1976, p. 127): The structure of the sentence can serve as a picture of the structure of the thought. In my view, this is not meant to imply that a thought has a unique structure, especially in the light of Frege's thesis that (structurally) different sentences can express the same thought. On the assumption that Frege regards (a) and (b) of the tentative contextual definition of "$N_x\varphi(x)$" as expressing the same (judgeable) content, the three decompositions that I described above display the structure of one and the same content each time in a different guise..

(ii) If the step of abstraction in Hume's Principle and in any other abstraction principle qua contextual definition of a term-forming operator were indeed intended to involve a decomposition of a sentence content in a mode different from the original mode, the new decomposition would have to be linked to a

non-standard-analysis of the right-hand side (a) of Hume's Principle that somehow leads to the left-hand side (b) and with it to the term-forming operator to be defined contextually. Hence, the original and the new way of decomposing the content would have to be strictly correlated to the original (standard) analysis of (a) and to a new, non-standard analysis of (a). Yet no matter how often we analyze (a) along Fregean lines, neither the cardinality operator, nor even "=" will miraculously emerge as a result of our analysis. And by way of decomposing the thought expressed by (a) in accordance with any analysis of (a), we shall never succeed in extracting the sense of the cardinality operator (or the sense of the predicate "*n* is a cardinal number"), nor even the sense of "=". In short, analyzing (a) in such a manner that it somehow effects the transition from (a) to (b) (from one structure of the thought under consideration to another), would be to make the impossible possible. The upshot is that Frege's way of describing abstraction in *Grundlagen*, §64 is marred by lacking precision and clarity. In particular, his talk of splitting up or dissecting a content in a way different from the original way is misguided. It gave rise to much misinterpretation in the literature, suggesting that abstraction à la Frege in *Grundlagen* is meant to be akin to the decomposition of a thought into thought components via gap formation, or in other words: that abstraction rests on the transition from one way of decomposing a thought to another..

(iii) Instead of characterizing the transition from (a) to (b) in the contextual definition of the cardinality operator as a recarving or dissection of a content in a way different from the original one[22], Frege (and presumably also Wright) should perhaps have said something like this: we have here the transition from one mode of speaking (a) to another (b) involving the desired function-name (singular term-forming operator) which (a) does not contain, or shorter: that one and the same content is presented in two different ways by (a) and (b). In Hume's Principle, the move from right to left, unlike that in a first-order abstraction principle, obviously involves stepping down from level two to level one.[23] Note also that by making the transition from (a) to (b) in Hume's Principle we do not directly obtain the concept of cardinal number, as Frege's way of speaking in §64 seems to suggest. What we do gain is the second-level

[22]When in *Grundlagen*, §64 Frege says that the judgement "The straight line *a* is parallel to the straight line *b*" can be construed as an equation, he chooses an infelicitous way of phrasing. This judgement qua sentence (note his use of quotation marks) can never be construed as an equation in a strict sense, although it can be *transformed* into an equation. The ensuing statement "We replace the symbol ∥ with the more general symbol =, by distributing the particular content of the former symbol to *a* and to *b*" is rather metaphorical and far from being clear. The linguistic operation of replacing a symbol with a more general symbol has nothing to do with abstraction. And what is it to mean precisely that the particular content of "=" is distributed between *a* and *b*? By the way, replacing the symbol "∥" by "=" in "*a* ∥ *b*" would be permissible at least from a syntactic point of view. The matter stands differently in the case of Hume's Principle. The sign for the second level relation of equinumerosity cannot literally be replaced by "=".

[23]According to Frege, identity is a first-level relation.

cardinality function, and it is in terms of this function that the first-level concept of cardinal number can subsequently be defined in second-order logic. Given Frege's awareness in *Grundlagen* that one and and the same judgeable content or thought can be decomposed in distinct ways, he certainly knew that a numerical equation such as "The number of planets = 9" (and the thought expressed by it) can be analyzed not only in the standard fashion, namely into "=" and the two singular terms flanking "=" (decomposed into the senses of these expressions), but also into the second-level predicate "the number of φs = 9" and the first-level predicate "planet" (decomposed into the senses of these predicates). According to this non-standard interpretation, our sentence expresses that the first-level concept *planet* falls in or under the second-level concept *the number of* φs = 9. Recalling Frege's principle of numerical predication (PNP) from *Grundlagen* (§46, §57) "An ascription of number contains a predication of a concept", we might also say that the sentence "The number of planets = 9", if we interpret it in a non-standard fashion, contains a predication of a (first-level) concept, just as the ascription of number "The number 9 belongs to the concept *planet*" (or "There are nine planets") is a second-order statement under its standard interpretation. It seems to me that only in such a case are we justified in speaking of a recarving or reanalysis of the same content with respect to one and the same sentence. But again, despite first appearances this has nothing to do with Fregean abstraction. So, I think that the widespread, but misleading talk of "recarving", "reanalysis" or "reconceptualization" with respect to Fregean abstraction has to be abandoned—the sooner the better.[24]

Frege's Proof of Hume's Principle

The Proof Sketch in Grundlagen

In *Grundlagen*, §73, Frege outlines roughly the proof of Hume's Principle from the explicit definition of the cardinality operator in §68. In doing this, he seems to rely tacitly on an abstraction principle that states for second-level concepts and their

[24]Parsons (1997, p. 270) seems to accept Wright's idea that Fregean abstraction effects a reconceptualization for the case of first-order principles. In those cases, he says, "it seems that what we are doing is simply individuating the objects we have in a coarser way, one might say carving up the domain, or a part of it, a little differently." This is very vague and thus far from being clearer than Wright's characterization. Fine (1998, p. 532) introduces the term "definition by reconceptualization" and says (but does not further explain) that it rests on the idea that new senses may emerge from a reanalysis of a given sense. He further claims that the idea derives from *Grundlagen*, §§63–64. However, as I have argued, Frege's attempted contextual definitions via abstraction cannot sensibly be described in such a way that new senses emerge from a reanalysis of a given sense.

extensions what in *Grundgesetze* Basic Law V states for first-level concepts and their extensions. In order to prove Hume's Principle, he has to show, according to the definition of the cardinality operator, that (1) if the concepts F and G are equinumerous, then the extension of the concept *equinumerous with the concept F* concides with the extension of the *concept equinumerous with the concept G*: $Eq_x(F(x), G(x)) \rightarrow Ext\varphi(Eq_x(\varphi(x), F(x))) = Ext\varphi(Eq_x(\varphi(x), G(x)))$ ("*Ext*" is an abbreviation for "extension"). That is to say: he has to prove that "under this hypothesis" the following two sentences hold generally: (2) $Eq_x(H(x), F(x)) \rightarrow Eq_x(H(x), G(x))$ and (3) $Eq_x(H(x), G(x)) \rightarrow Eq_x(H(x), F(x))$. Thus, Frege is converting here the statement that the extensions of two specific second-level concepts are identical into the statement that these concepts are coextensive: $Ext\varphi(Eq_x(\varphi(x), F(x))) = Ext\varphi(Eq_x(\varphi(x), G(x))) \leftrightarrow \forall\varphi(Eq_x(\varphi(x), F(x)) \leftrightarrow Eq_x(\varphi(x), G(x)))$.[25] We might thus presume that in introducing extensions of second-level concepts he had in mind the following third-order abstraction principle:

$$Extf(M_\beta(f(\beta))) = Extf(N_\beta(f(\beta))) \leftrightarrow \forall f(M_\beta(f(\beta)) \leftrightarrow N_\beta(f(\beta))).$$

Yet, if this applies, it would remain obscure why Frege does not explicitly invoke this higher-order principle when he comes to introduce extensions of second-level concepts. As against this, someone might wish to argue that in *Grundlagen* Frege must have been aware that it is imperative to lay down such a principle as an axiom of a formal theory whose first-order domain comprises extensions of second-level concepts. Be this as it may, it remains true that in *Grundlagen* it is exclusively Hume's Principle that determines explicitly the identity conditions for cardinal numbers.[26]

Following Frege, (2) amounts to this: there is a relation T that correlates one-to-one the objects falling under H with those falling under G if there is a relation R that correlates one-to-one the objects falling under F with those falling under G, and if there is a relation S that correlates one-to-one the objects falling under H with those falling under F:

[25]See in this respect Schirn (2003, p. 213) and Heck (1995, p. 130).

[26]Note that the third-order abstraction principle above applies in *Grundlagen* only to second-level *concepts* (and their extensions), not to monadic second-level functions in general. I assume that around 1884 Frege did not yet consider extensions of concepts and extensions of relations as special cases of what after 1891 he calls value-ranges of one-place functions and value ranges of two-place functions (double value-ranges). Thanks to his device of "level-reduction", he is able to confine himself to the introduction of value-ranges of *first*-level functions. First-level functions which appear as arguments of second-level functions are represented by their value-ranges, "though of course not in such a way that they give up their places to them, for that is impossible" (Frege 1893, §34).

$$
\begin{aligned}
&[\exists R(\forall x(F(x) \rightarrow \exists y(R(x,y) \wedge G(y))) \wedge \forall y(\mathrm{G}(y) \rightarrow \exists x(R(x,y) \wedge F(x)))\\
&\quad \wedge \forall x \forall y \forall z((R(x,y) \wedge R(x,z) \rightarrow y = z) \wedge (R(x,z) \wedge R(y,z) \rightarrow x = y)))\\
&\quad \wedge \exists S(\forall x H(x) \rightarrow \exists y(S(x,y) \wedge F(y))) \wedge \forall y(F(y)\\
&\quad \rightarrow \exists x(S(x,y) \wedge H(x))) \wedge \forall x \forall y \forall z((S(x,y)\\
&\quad \wedge S(x,z) \rightarrow y = z) \wedge\ (S(x,z) \wedge S(y,z) \rightarrow x = y)))]\\
&\quad \rightarrow \exists T(\forall x H(x) \rightarrow \exists y(T(x,y) \wedge G(y))) \wedge \forall y(G(\mathrm{y})\\
&\quad \rightarrow \exists x(T(x,y) \wedge H(x))) \wedge \forall x \forall y \forall z((T(x,y) \wedge T(x,z) \rightarrow y = z)\\
&\quad \wedge (T(x,z) \wedge T(y,z) \rightarrow x = y))).
\end{aligned}
$$

Frege goes on to say that such a relation T can in fact be given; it lies in the content "There exists an object to which c stands in the relation S and which stands to b in the relation R"

$$\exists x(S(c,x) \wedge R(x,b)),$$

if we detach from it c and b (considered to be relation-points). He further claims that one can show that the (dyadic) relation $\exists x(S(z,x) \wedge R(x,y))$ correlates one-to-one the objects falling under H with those falling under G. He concludes this proof sketch by saying that (3) can be proved in a similar way and seems to suggest in a footnote to §73 that the converse of (1), namely

$$(4) \quad Ext\varphi(Eq_x(\varphi(x), F(x))) = Ext\varphi(Eq_x(\varphi(x), G(x))) \rightarrow Eq_x(F(x), G(x))$$

can be proved in a similar fashion as (1).

The Formal Proof in Grundgesetze

Throughout my account of Frege's proof of Hume's Principle in *Grundgesetze* I transcribe Frege's concept-script formulae largely into "modern" notation. In particular, I replace Frege's symbol for negation by "¬", his symbol for the conditional function by "→", his sign for the universal quantifier by "∀", and the symbol for the value-range operator by "VR". As to the role of "=" in Frege's concept-script, we must bear in mind his assimilation of declarative sentences to proper names. Due to the fact that he construes referential expressions with the syntactic structure of declarative sentences as proper names referring either to the True or to the False, he feels entitled to employ "=" not only between ordinary singular terms, but also between declarative sentences (truth-value names); Basic Law V is a paradigm case. In my transcriptions, I follow Frege in this respect, although "=", when flanked by two truth-value names, could be replaced by "↔" without essentially affecting the content of a given formula.

Instead of Frege's symbol for the "membership-function" that resembles the sign for the operation of intersection in set theory, I employ "$\in$". I am of course aware that the meaning usually attached to "$\in$" does not coincide with the meaning that Frege assigns to his symbol, but the meaning of the latter is akin to the meaning of the former. Again, this should not cause any problem, especially since below I shall list Frege's definition of the membership-function in modern notation. Thus, whenever the symbol "$\in$" occurs in my transcription of a Fregean formula—and it does of course occur quite often—it must always be understood exactly in the sense that the definition bestows upon it. As a matter of fact, Frege employs only negation, implication, the universal quantifier and identity from the repertoire of logical signs available in standard first- and second-order logic.

Frege refers to the sign that he uses for the first-order and second-order universal quantifier as "concavity" ("*Höhlung*"). Since I replace the concavity everywhere by "$\forall$", the word "concavity" has accordingly to be replaced in my text by "universal quantifier" or "$\forall$". However, I shall retain Frege's use of German letters for bound variables throughout my exposition. Thus, instead of "$\smile\!\!\frown\!\!\smile\ \mathfrak{a} = \mathfrak{a}$", for example, I write "$\forall\mathfrak{a}(\mathfrak{a} = \mathfrak{a})$". I shall also retain Frege's choice of Roman object- and function-letters so that formulae in their original concept-script notation can be compared more conveniently with the corresponding ones in my transcription. Finally, I shall retain Frege's use of small Greek vowels, except for his use of the smooth breathing which he employs for his value-range notation. Instead of "$\grave{\varepsilon}\grave{\alpha}(\varepsilon = \alpha)$", for example, I write "VRαVR$\varepsilon(\varepsilon = \alpha)$", where "VR" is supposed to correspond to the smooth breathing.

Finally, a word about Frege's judgement stroke and the horizontal. In my transcription of the sentences that he puts forward or proves in *Grundgesetze*, I dispense with the judgement stroke, which, for obvious reasons, appears in his symbolism always in tandem with the horizontal. I also dispense with the horizontal with a very few exceptions. From a logical point of view, the horizontal is dispensable in Frege's logical theory. The concept —ξ can be reduced to the relation $\xi = \zeta$, since —ξ is coextensive with $\xi = (\xi = \xi)$. However, without the notational benefit that Frege derives from "—ξ", his two-dimensional concept-script would not even have got off the ground.

Since I could not typographically reproduce some of the simple signs for complex functions that Frege introduces via definitions—in particular, the signs for the designation of the functions *composite relation*, *converse of a relation* and *cardinal number of a concept*—I have replaced them by more accessible signs.

I what follows, I first list the definitions on which Frege relies in his formal proof of Hume's Principle and then state the rules to which he likewise appeals.

Definitions

The relation of an object falling within the extension of a concept (*Beziehung des Hineinfallens eines Gegenstandes in einen Begriffsumfang*), §34 (Def. A):

$$a \in u := \backslash \mathrm{VR}\alpha(\neg\forall\mathfrak{g}(u = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = \alpha))$$

Composite relation (*zusammengesetzte Beziehung*), §54 (Def. B):

$$p|q := \mathrm{VR}\alpha\mathrm{VR}\varepsilon(\neg\forall\mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg\varepsilon \in (\mathfrak{r} \in p)))$$

Single-valuedness of a relation (*Eindeutigkeit einer Beziehung*), §37 (Def. Γ):

$$\mathrm{I}p := \forall\mathfrak{e}\forall\mathfrak{d}(\mathfrak{e} \in (\mathfrak{d} \in p) \rightarrow \forall\mathfrak{a}(\mathfrak{e} \in (\mathfrak{a} \in p) \rightarrow \mathfrak{d} = \mathfrak{a}))$$

Mapping-into by a relation (*Abbildung durch eine Beziehung*), §38 (Def. Δ):

$$\rangle p := \mathrm{VR}\alpha\mathrm{VR}\varepsilon(\neg(\mathrm{I}p \rightarrow \neg\forall\mathfrak{d}(\forall\mathfrak{a}(\mathfrak{d} \in (\mathfrak{a} \in p) \rightarrow \neg\mathfrak{a} \in \alpha) \rightarrow \neg\mathfrak{d} \in \varepsilon)))$$

Converse of a relation (*Umkehrung einer Beziehung*), §39 (Def. E):

$$\int p := \mathrm{VR}\alpha\mathrm{VR}\varepsilon(\alpha \in (\varepsilon \in p))$$

The cardinal number of a concept (*die Anzahl eines Begriffs*), §40 (Def. Z); I employ "N" instead of Frege's sign for the cardinality operator:

$$\mathrm{N}u := \mathrm{VR}\varepsilon(\neg\forall\mathfrak{q}(u \in (\varepsilon \in \rangle\int\mathfrak{q}) \rightarrow \neg\varepsilon \in (u \in \rangle\mathfrak{q})))$$

Rules

Since I replace Frege's vertical conditional-stroke as part of his two-dimensional concept-script everywhere by the symbol "→", his term "subcomponent" ("*Unterglied*") must be replaced everywhere by "antecedent" and "supercomponent" ("*Oberglied*") by "consequent." For the reason that I mentioned above, the word "smooth breathing" does not occur in the modified formulation of the rules..

3. *Contraposition*

An antecedent in a sentence may be permuted with a consequent, if one also inverts their truth-values..

4. *Fusion of equal antecedents*

An antecedent that occurs repeatedly in the sane sentence only needs to be written once..

5. *Transformation of a Roman letter into a German letter*

A Roman letter may be replaced wherever it occurs in a sentence by the same German letter, namely an object-letter by an object letter and a function-letter by a function-letter. At the same time, the latter must be placed after a universal quantifier in front of a consequent, outside of which the Roman letter did not occur. If in this consequent the scope of a German letter is completely contained and the

Roman letter occurs within this scope, then the German letter that is to be introduced for the latter must be distinct from the former..

7. *Inferring* (b)

If the same combination of signs (either a proper name or a Roman object-marker) occurs in one sentence as consequent and in another as antecedent, then a sentence may be inferred in which the consequent of the second sentence appears as a consequent and all antecedents of both, save that mentioned, appear as antecedents. Equal antecedents may here be fused according to rule (4).

Transition-signs: ( ) : — — —
and ( ) : : — — — Combined inferences thus:
(,) : : = = = and (,) : : ———— (here I had to change the last component of the sign in Frege's notation).
— — — —.

9. *Citing sentences*: *replacement of Roman letters*

When citing a sentence by its label one may effect a simple inference by uniformly replacing a Roman letter within the sentence by the same proper name or the same Roman object-marker.

Likewise, one may replace all occurrences in a sentence of a Roman function-letter, '*f*', '*g*', '*h*', '*F*', '*G*', '*H*' (cf. Frege 1893, §§19–20) by the same name or Roman marker of a one-place or two-place first-level function, depending on whether the Roman letter indicates a one-place or two-place function.

When citing Basic Law (IIb)—"$\forall \mathfrak{f}(M_\beta(M_\beta(\mathfrak{f}(\beta))) \rightarrow M_\beta(f\beta))$"—(cf. §25)—one may replace both occurrences of 'M_β' in it by the same name or Roman marker of a second-level function with one argument of the second kind (that is, with a monadic first-level function).

As to the question which argument-places are to be regarded as related, the following two rules must be observed:

All occurrences of a German letter within its scope, save those within an enclosed scope of the same letter or after "∀", are *related* argument-places of the corresponding function.

All occurrences of a small Greek vowel within its scope, save those within an enclosed scope of the same letter or together with the value-range sign "VR" are likewise *related* argument-places of the corresponding function..

11. *Citing sentences*; *replacement of Greek vowels*

When citing a sentence by its label, one may uniformly replace a German letter after "VR" and at all argument-places of the corresponding function[27] by one and the

[27]Frege uses "function" here and I have respected his choice. However, strictly speaking, it is only the function-name that has an argument-place or argument-places that are suitable to take names of the appropriate syntactic category. Plainly, in a function one cannot replace a letter by another one.

same distinct letter, that is, an object letter by an object-letter, and a function-letter by a function-letter, if no German letter occurring in a scope within its own thereby becomes the same as the one whose scope is enclosed.

Hume's Principle

In Frege (1893, §53), Hume's Principle is introduced as follows:

> The cardinal number of a concept is equal to the cardinal number of a second concept, if a relation maps the first into the second, and if the converse of this relation maps the second into the first.

By using "→" for the designation of Frege's conditional function, "∈" for the designation of his membership function, "⟩" for the designation of the mappping-into by a relation, "∫" for the designation of the converse of a relation and "Nξ" for the designation of the first-level cardinality function (§40) we can write Hume's Principle as it appears in *Grundgesetze* in symbolic notation as follows:

$$v \in (u \in\rangle \int q) \rightarrow (u \in (v \in\rangle \mathrm{q}) \rightarrow \mathrm{N}u = \mathrm{N}v)$$

In equivalent notation:

$$(v \in (u \in\rangle \int q) \wedge u \in (v \in\rangle \mathrm{q})) \rightarrow \mathrm{N}u = \mathrm{N}v$$

We saw that as early as in *Grundlagen* Frege was thinking about the proof of Hume's Principle. This does not imply that at that time he already surveyed exactly all the steps to be carried out formally in his concept-script. We must also bear in mind that in 1893 the original concept-script of 1879, that basically underlies Frege's informal considerations in *Grundlagen*, had undergone a number of internal and partly thoroughgoing changes due to a far-reaching development of Frege's logical views. Since the proof of Hume's Principle is of paramount importance for Frege's logical foundation of cardinal arithmetic, it is worth seeing how in Frege (1893) the proof gets off the ground and how it proceeds.

The proof is divided into (a) the proof of

$$\text{“}u \in (v \in\rangle q) \rightarrow (w \in (u \in\rangle p) \rightarrow w \in (v \in\rangle q|p))\text{”}(\S\S 54{-}59)$$

and (b) the proof of

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \rightarrow (u \in (v \in\rangle q) \rightarrow (\neg\forall \mathrm{q}(u \in (\mathrm{w} \in\rangle \int \mathrm{q}) \rightarrow \neg \mathrm{w} \in (u \in\rangle \mathrm{q}))\\ &\quad \rightarrow (\neg\forall \mathrm{q}(v \in (\mathrm{w} \in\rangle \int \mathrm{q}) \rightarrow \neg \mathrm{w} \in (v \in\rangle \mathrm{q}))))\text{”}(\S\S 60-61)\end{aligned}$$

It goes on in §63 with the derivation of

$$(27) \quad \text{“I}\smile\!\frown q \rightarrow (u \in (v \in\rangle q) \rightarrow u \in (v \in\rangle \smile\!\frown q)\text{”}$$

and is then completed in §65 by arriving finally at Hume's Principle (sentence 32).

The proof as presented by Frege proceeds in six stages of what he calls *construction* (*Aufbau*). Given its length and complexity, I assume that it did not fall into his lap. Each construction (see §§55, 57, 59, 61, 63, 65) is preceded by what Frege terms "analysis" (*Zerlegung*). The force of the proof is of course to be sought only under the heading "construction". For reasons of space, I confine myself to considering only the first, the fifth and the final construction in the entire proof. However, I shall reconstruct every analysis 1–6.

Analysis 1 (§54)

From Frege's very sketchy outline in *Grundlagen* of the first steps to be carried out in the envisaged formal proof of Hume's Principle, it is clear that in *Grundgesetz*e, §54 he begins by stating that according to his definition (Z) of the (first-order) cardinality operator Hume's Principle

$$\text{“}v \in (u \in\rangle \smile\!\frown q) \rightarrow (u \in (v \in\rangle q) \rightarrow \mathrm{N}u = \mathrm{N}v\text{”} \qquad (\alpha)$$

is a consequence of

$$\begin{aligned}&\text{“}v \in (u \in\rangle \smile\!\frown q) \rightarrow (u \in (v \in\rangle q) \rightarrow \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(u \in (\varepsilon \in\rangle \smile\!\frown q)\\ &\quad \rightarrow \neg\varepsilon \in (u \in\rangle \mathfrak{q}))) = \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(v \in (\varepsilon \in\rangle \smile\!\frown \mathfrak{q}) \rightarrow \neg\varepsilon \in (v \in\rangle \mathfrak{q}))))\text{”}\end{aligned} \qquad (\beta)$$

Plainly, (β) corresponds to the sentence from Frege (1884, §73) that I rendered in symbolic notation as follows:

$$Eq_x(F(x), G(x)) \rightarrow Ext\varphi(Eq_x(\varphi(x), F(x))) = Ext\varphi(Eq_x(\varphi(x), G(x))).$$

Now, (β) must be derived—by appeal to (Va)

$$\text{“}\forall \mathfrak{a}(f(\mathfrak{a}) = g(\mathfrak{a})) \rightarrow F(\mathrm{VR}\varepsilon(f(\varepsilon)) = F(\mathrm{VR}\alpha(\mathfrak{g}(\alpha))\text{”},$$

and by applying rule (5) (=the transformation of a Roman letter into a German letter)—from

$$\begin{aligned}&\text{“}v \in (u \in\rangle \smile\!\frown q) \rightarrow (u \in (v \in\rangle q) \rightarrow (\neg\forall \mathfrak{q}(u \in (w \in\rangle \smile\!\frown \mathfrak{q})\\ &\quad \rightarrow \neg w \in (u \in\rangle \mathfrak{q}))) = (\neg\forall \mathfrak{q}(v \in (w \in\rangle \smile\!\frown \mathfrak{q}) \rightarrow \neg w \in (v \in\rangle \mathfrak{q}))))\text{”}\end{aligned} \qquad (\gamma)$$

which has to be derived using (IVa):

$$\text{“}(b \to a) \to ((a \to b) \to (\text{—}a) = (\text{—}b))\text{”}$$

To do this, Frege needs the sentences

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \to (u \in (v \in\rangle q) \to \neg\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q})\\ &\quad \to \neg w \in (v \in\rangle \mathfrak{q})) \to (\neg\forall \mathfrak{q}(u \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (u \in\rangle \mathfrak{q}))))\text{”}\end{aligned} \tag{δ}$$

and

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \to (u \in (v \in\rangle q) \to (\neg\forall \mathfrak{q}(u \in (w \in\rangle \int \mathfrak{q})\\ &\quad \to \neg w \in (u \in\rangle \mathfrak{q})) \to \neg\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (v \in\rangle \mathfrak{q}))))\text{”}\end{aligned} \tag{ϵ}$$

which differ only with respect to the use of “v” and “u” in their quantified components.

Frege suggests to interchange “u” with “v” and to replace “q” by “$\int q$” in (ε) which results in

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \to (u \in (v \in\rangle q) \to (\neg\forall \mathfrak{q}(u \in (w \in\rangle \int \mathfrak{q})\\ &\quad \to \neg w \in (u \in\rangle \mathfrak{q})) \to \forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (v \in\rangle \mathfrak{q}))))\text{”}\end{aligned} \tag{ζ}$$

In oder to derive (δ) from (ζ) by rule (7), he requires

$$\text{“}u \in (v \in\rangle q) \to u \in (v \in\rangle \int q)\text{”}. \tag{η}$$

(ε) has to be proved first. It results by contraposition from

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \to (u \in (v \in\rangle q) \to (\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q})\\ &\quad \to \neg w \in (v \in\rangle \mathfrak{q})) \to \forall \mathfrak{q}(u \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (u \in\rangle \mathfrak{q}))))\text{”},\end{aligned} \tag{θ}$$

which in turn follows, according to rule (5), from

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \to (u \in (v \in\rangle q) \to (\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q})\\ &\quad \to \neg w \in (v \in\rangle \mathfrak{q})) \to (u \in (w \in\rangle \int p) \to \neg w \in (u \in p))))\text{”}.\end{aligned} \tag{ι}$$

In order to facilitate the grasp of the thought expressed by (ι), Frege proposes to transform it by contraposition into

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \to (u \in (v \in\rangle q) \to (w \in (u \in\rangle p) \to (u \in (w \in\rangle \int p)\\ &\quad \to \neg\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (v \in\rangle \mathfrak{q})))))\text{”}.\end{aligned} \tag{κ}$$

For the sake of convenience, he now says “u-concept” instead of “concept whose extension is indicated by ‘u’”, “p-relation” instead of “relation whose extension is

indicated by p", "the p-relation maps the w-concept into the u-concept" instead of "the objects falling under the w-concept are correlated single-valuedly with the objects falling under the u-concept by the p-relation". (κ) can now be put in words as follows:

"If the converse of the p-relation maps the u-concept into the w-concept and the p-relation maps the w-concept into the u-concept, and if further the q-relation maps the u-concept into the v-concept and the converse of the q-relation maps the v-concept into the u-concept, then there is a relation that maps the w-concept into the v-concept and its converse maps the v-concept into the w-concept."

Such a relation is clearly one that is composed from the p-relation and the q-relation (cf. also *Grundlagen*, §73). Frege accordingly goes on to set up the definition of *composite relation* (Def. B) which I listed above. What is required for the proof of Hume's Principle is now the proof of sentence (λ) "If the p-relation maps the w-concept into the u-concept and if the q-relation maps the u-concept into the v-concept, then the $p|q$-relation that is composed from the two maps the w-concept into the v-concept", to which I referred earlier as Frege's goal of part (a) of the proof of Hume's Principle.

$$\text{“}u \in (v \in\rangle q) \rightarrow (w \in (u \in\rangle p) \rightarrow w \in (v \in\rangle (q|p)))\text{”} \qquad (\lambda)$$

To carry out the proof of (λ), Frege needs

$$\text{“}v \in (u \in\rangle \int q) \rightarrow (u \in (w \in\rangle \int p) \rightarrow v \in (w \in\rangle \int (p|q)))\text{”} \qquad (\mu)$$

which can be reduced to (λ) with the aid of

$$\text{“}\int (p|q) = \int q|\int p)\text{”} \qquad (\nu)$$

Turning our attention to the definition of mapping-into by a relation (cf. Def. Δ above), we see that the following two sentences must be proved:

$$\begin{aligned}&\text{“}u \in (v \in\rangle q) \rightarrow (w \in (u \in\rangle p)\\ &\quad \rightarrow \forall \mathfrak{d}(\forall \mathfrak{a}(\mathfrak{d} \in (\mathfrak{a} \in (p|q)) \rightarrow \neg \mathfrak{a} \in \mathrm{v}) \rightarrow \neg \mathfrak{d} \in w)\text{”}\xi)\end{aligned}$$

and

$$\text{“I}p \rightarrow \text{I}q \rightarrow \text{I}(p|q)\text{”} \qquad (\mathrm{o})$$

We obtain (ξ) from

$$\begin{aligned}&\text{“}u \in (v \in\rangle q) \rightarrow (w \in (u \in\rangle p)\\ &\quad \rightarrow \forall \mathfrak{a}(d \in (\mathfrak{a} \in (p|q)) \rightarrow \neg \mathfrak{a} \in \mathrm{v}) \rightarrow \neg d \in w))\text{”}\end{aligned} \qquad (\pi)$$

by applying rule (5).

In order to render the thought expressed by (π) conveniently in words, Frege applies again contraposition as before in a similar case. The resulting sentence

$$\begin{aligned}&\text{“}u \in (v \in\rangle q) \rightarrow (w \in (u \in\rangle p)\\ &\quad \rightarrow (d \in w \rightarrow \neg\forall\mathfrak{a}(d \in (\mathfrak{a} \in (p|q)) \rightarrow \neg\mathfrak{a} \in \mathrm{v})))\text{”}\end{aligned} \tag{ρ}$$

says:

“If d (=the object which is indicated by “d”) falls under the w-concept, and if the w-concept is mapped into the u-concept by the p-relation, and if the u-concept is mapped into the v-concept by the q-relation, then there is an object that falls under the v-concept and to which d stands in the $(p \mid q)$-relation.”

Frege mentions that the proof will rely on

$$\text{“}d \in (e \in p) \rightarrow (\mathrm{e} \in (m \in q) \rightarrow d \in (m \in (p|q)))\text{”} \tag{σ}$$

in words: “If d stands in the p-relation to e, and if e stands in the q-relation to m, then d stands in the $p \mid q$-relation to m.”

(σ) needs to be derived from

$$\text{“}(\neg\forall\mathfrak{r}(\mathfrak{r} \in (m \in q) \rightarrow \neg\mathfrak{a} \in (\mathfrak{r} \in p))) = d \in (m \in (p|q))\text{”} \tag{τ}$$

which is a consequence of Def. (B). In order to prove (τ), Frege requires

$$\text{“}f(a,b) = a \in (\mathfrak{b} \in \mathrm{VR}\alpha\mathrm{VR}\varepsilon(f(\varepsilon,\alpha)))\text{”}. \tag{υ}$$

(υ) must be reduced to

$$\text{“}f(a) = a \in \mathrm{VR}\varepsilon(f(\varepsilon))\text{”}, \tag{φ}$$

which is to be derived from the definition of the relation of an object falling within the extension of a concept (Def. A). According to Def. (A), Frege must prove

$$\text{“}f(a) = \backslash\mathrm{VR}\alpha(\neg\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon) = \neg\mathfrak{g}(a) = \alpha))\text{”} \tag{χ}$$

by means of (VI)

$$\text{“}\forall\mathfrak{a}(f(\mathfrak{a}) = (a = \mathfrak{a})) \rightarrow a = \backslash\mathrm{VR}\varepsilon(f(\varepsilon))\text{”}$$

and

$$\text{“}\forall\mathfrak{a}((\neg\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = \mathfrak{a}) = (f(a) = \mathfrak{a}))\text{”} \tag{ψ}$$

taking

$$\text{“}(\neg\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = \xi)\text{”}$$

for “$f(\xi)$” in (VIa) and replacing “a” by “$f(a)$”. (ψ) results by rule (5) from

$$\text{“}(\neg\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = b)) \rightarrow (f(a) = b)\text{”} \qquad (\omega)$$

which has to be proved by appeal to (IVa). To do this, Frege needs

$$\text{“}f(a) = b \rightarrow \neg\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = b)\text{”} \qquad (\alpha')$$

and

$$\text{“}\neg\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = b) \rightarrow f(a) = b\text{”}. \qquad (\beta')$$

(α′) follows by contraposition from

$$\text{“}\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = b) \rightarrow f(a) = b\text{”} \qquad (\gamma')$$

If we now write (IIb)

$$\text{“}\forall\mathfrak{f}(M_\beta(\mathfrak{f}(\beta))) \rightarrow M_\beta(f(\beta))\text{”}$$

in the form

$$\begin{aligned}&\text{“}\forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = b)\\ &\rightarrow (\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(f(\varepsilon)) \rightarrow \neg f(a) = b)\text{”},\end{aligned}$$

we see that (γ′) follows from it and (IIIe)

$$a = a$$

(β′) follows by contraposition from

$$\text{“}\neg f(a) = b \rightarrow \forall\mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon))(\varepsilon)) \rightarrow \neg\mathfrak{g}(a) = b)\text{”} \qquad (\delta')$$

and this follows in turn by rule (5) from

$$\text{“}\neg f(a) = b \rightarrow (\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(g(\varepsilon)) \rightarrow \neg g(a) = b)\text{”}. \qquad (\epsilon')$$

We obtain this formula by applying rule (7) and by invoking (Vb)

$$\text{“}\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\alpha(\mathrm{g}(\alpha)) \rightarrow f(a) = g(b)\text{”}$$

from

$$\text{“}\neg f(a) = b \rightarrow (f(a) = g(a)) \rightarrow \neg g(a) = b\text{”}$$

which, by permutation of antecedents, is just a special case of (IIIc): “$a = b \rightarrow (f(a) = f(b))$”. So much for the details of the first analysis. Let us now turn to the first construction.

Construction 1 (§55)

$$(\mathrm{Vb})\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\alpha(\mathrm{g}(\alpha)) \rightarrow f(a) = g(b)$$

(IIIc): – – – – – – – – – –

$$f(a) = b \rightarrow \mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathrm{g}(\varepsilon)) \rightarrow \neg g(a) = b \qquad (\alpha)$$

$$f(a) = b \rightarrow \forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b) \qquad (\beta)$$

$$\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b) \rightarrow f(a) = b \qquad (\gamma)$$

(IVa):———————

$$\begin{aligned} &f(a) = b \rightarrow \neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b) \\ &\quad \rightarrow (\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b)) \rightarrow (\text{—}f(a) = b) \end{aligned} \qquad (\delta)$$

——————— ♦ ———————

IIIe $\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(f(\varepsilon))$

(IIb): ———————–

$$\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b) \rightarrow \neg f(a) = b \qquad (\epsilon)$$

$$\times$$

$$f(a) = b\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b) \qquad (\zeta)$$

(δ): ———————

$$\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b) = (\text{—}f(a) = b) \qquad (\eta)$$

(IIIa) : ———————

$$\begin{aligned} (\text{—}f(a) = b) = f(a) = b) \rightarrow ((\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) \\ = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b)) = (f(a) = b) \end{aligned} \qquad (\theta)$$

(IIIi) :: ———————

$$(\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = b)) = (f(a) = b) \qquad (\iota)$$

$$\div$$

$$\forall \mathfrak{a}(\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = \mathfrak{a})) = (f(a) = \mathfrak{a}) \qquad (\kappa)$$

(VIa): ———————

$$f(a) = \backslash\mathrm{VR}\alpha(\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = \alpha)) \qquad (\lambda)$$

(IIIa): ———————

$$\begin{gathered}\backslash\mathrm{VR}\alpha(\neg\forall \mathfrak{g}(\mathrm{VR}\varepsilon(f(\varepsilon)) = \mathrm{VR}\varepsilon(\mathfrak{g}(\varepsilon)) \rightarrow \neg \mathfrak{g}(a) = \alpha)) \\ = a \in \mathrm{VR}\varepsilon(f(\varepsilon)) \rightarrow f(a) = a \in \mathrm{VR}\varepsilon(f(\varepsilon))\end{gathered} \qquad (\mu)$$

(A) :: ———————

$$f(a) = a \in \mathrm{VR}\varepsilon(f(\varepsilon)) \qquad (1)$$

———— ♦ ————

$$1 \quad f(a,b) = a \in \mathrm{VR}\varepsilon(f(\varepsilon, b))$$

(IIIc) : ———————

$$\begin{gathered}\mathrm{VR}\varepsilon(f(\varepsilon, b)) = b \in \mathrm{VR}\alpha\mathrm{VR}\varepsilon(f(\varepsilon, \alpha)) \\ \rightarrow f(a,b) = a \in (b \in \mathrm{VR}\alpha\mathrm{VR}\varepsilon(f(\varepsilon, \alpha))\end{gathered} \qquad (\alpha)$$

1 :: ———————

$$f(a,b) = a \in \mathrm{VR}\varepsilon(f(\varepsilon, b)) \qquad (2)$$

$$(\mathrm{IIIc}): \mathrm{VR}\alpha\mathrm{VR}\varepsilon(f(\varepsilon, \alpha)) = q \rightarrow f(a,b) = a \in (b \in q) \qquad (3)$$

———— ♦ ————

$$\mathrm{VR}\alpha\mathrm{VR}\varepsilon(\neg\forall \mathfrak{r}(\mathfrak{r} \in (\alpha \in q) \rightarrow \neg\varepsilon \in (\mathfrak{r} \in p))) = p|q \qquad (\mathrm{B})$$

(3) : ———————

$$(\neg\forall \mathfrak{r}(\mathfrak{r} \in (m \in q) \rightarrow \neg d \in (\mathfrak{r} \in p))) = d \in (m \in (p|q)) \qquad (4)$$

(IIIc) : ———————

$$(\neg\forall \mathfrak{r}(\mathfrak{r} \in (m \in q) \rightarrow \neg d \in (\mathfrak{r} \in p)) \rightarrow d \in (m \in (p|q)) \qquad (\alpha)$$

$$\times$$

$$\neg d \in (m \in (p|q)) \rightarrow \neg\forall\mathfrak{r}(\mathfrak{r} \in (m \in q) \rightarrow \neg d \in (\mathfrak{r} \in p)) \qquad (\beta)$$

(IIa) : – – – – – – – – –

$$\neg d \in (m \in (p|q)) \rightarrow (e \in (m \in q) \rightarrow d \in (e \in p)) \qquad (\gamma)$$

$$\times$$

$$d \in (e \in p)) \rightarrow (e \in (m \in q) \rightarrow d \in (m \in (p|q) \qquad (5)$$

Analysis 2 (§56)
As Frege points out, the proof of

$$\begin{array}{l} \text{“}u \in (v \in\rangle q) \rightarrow (w \in (u \in\rangle p) \rightarrow \forall\mathfrak{a}(d \in (\mathfrak{a} \in (p|q)) \\ \quad \rightarrow \neg\mathfrak{a} \in \mathrm{v}) \rightarrow \neg d \in w))\text{”}(\S 54, \pi) \end{array} \qquad (\alpha)$$

requires to go back to the definition of *mapping-into by a relation* (Def. Δ). From this he derives

$$\text{“}w \in (u \in\rangle p) \rightarrow \forall\mathfrak{a}(d \in ((\mathfrak{a} \in p) \rightarrow \neg\mathfrak{a} \in u)\neg d \in w))\text{”} \qquad (\beta)$$

In order to reach (α) from this, Frege needs the formula

$$\begin{array}{l} \text{“}\forall\mathfrak{a}(d \in (\mathfrak{a} \in p|q) \rightarrow \neg\mathfrak{a} \in v) \rightarrow (u \in (v \in\rangle q) \\ \quad \rightarrow \forall\mathfrak{a}(d \in (\mathfrak{a} \in p) \rightarrow \neg\mathfrak{a} \in u))\text{”} \end{array} \qquad (\gamma)$$

which he obtains by rule (5) from

$$\begin{array}{l} \text{“}\forall\mathfrak{a}(d \in (\mathfrak{a} \in p|q) \rightarrow \neg\mathfrak{a} \in v) \rightarrow (u \in (v \in\rangle q) \\ \quad \rightarrow (d \in (e \in p) \rightarrow \neg e \in u))\text{”}. \end{array} \qquad (\delta)$$

Now, (β) can be transformed into

$$\text{“}u \in (v \in\rangle p) \rightarrow (\forall\mathfrak{a}(e \in (\mathfrak{a} \in q) \rightarrow \neg\mathfrak{a} \in v)\neg e \in u))\text{”} \qquad (\beta)$$

In order to obtain (δ) from this, Frege needs the formula

$$\begin{aligned}&\text{“}\forall\mathfrak{a}(d \in (\mathfrak{a} \in p|q) \rightarrow \neg\mathfrak{a} \in v) \rightarrow (d \in (e \in p)\\&\quad \rightarrow \forall\mathfrak{a}(e \in (\mathfrak{a} \in q) \rightarrow \neg\mathfrak{a} \in v))\text{”}\end{aligned} \tag{ϵ}$$

which follows by rule (5) from

$$\begin{aligned}&\text{“}\forall\mathfrak{a}(d \in (\mathfrak{a} \in p|q) \rightarrow \neg\mathfrak{a} \in v) \rightarrow (d \in (e \in p)\\&\quad \rightarrow (e \in (m \in q) \rightarrow \neg m \in v))\text{”},\end{aligned} \tag{ζ}$$

which in turn can easily proved by appeal to (IIa) “$\forall\mathfrak{a}(f(\mathfrak{a})) \rightarrow f(a)$” and (5).

Hence, what matters is the derivation of (β) from (Δ). This can be accomplished by drawing on

$$\text{“VR}\alpha\text{VR}\varepsilon(f(\varepsilon, \alpha)) = q \rightarrow F(a \in (b \in q)) \rightarrow Ff(a, b)\text{”} \tag{η}$$

which follows from (3).

Analysis 3 (§58)

After having carried out the construction 2 according to the details mentioned in analysis 2, Frege turns to analysis 3. He begins by stating that he must now prove that the relation that is composed from the *p*-relation and the *q*-relation is single-valued if both the *p*-relation and the *q*-relation are single-valued, in symbols:

$$\text{“I}p \rightarrow (\text{I}q \rightarrow \text{I(p|q)})\text{”} \tag{α}$$

According to Def. (Γ), it must be proved

$$\begin{aligned}&\text{“I}p \rightarrow (\text{I}q \rightarrow \text{I}(p|q)) \rightarrow (\forall\mathfrak{e}\forall\mathfrak{d}(\mathfrak{e} \in (\mathfrak{d} \in (p|q))\\&\quad \rightarrow \forall\mathfrak{a}(\mathfrak{e} \in (\mathfrak{a} \in (p|q)) \rightarrow \mathfrak{d} = \mathfrak{a})))\text{”}\end{aligned} \tag{β}$$

which according to rule (5) results from

$$\text{“I}p \rightarrow (\text{I}q \rightarrow (e \in (d \in (p|q)) \rightarrow (e \in (a \in (p|q)) \rightarrow d = a)))\text{”}. \tag{γ}$$

From the definition of *composite relation* (Def. B) one can easily derive

$$\text{“}e \in (a \in (p|q)) \rightarrow \neg\forall\mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg e \in (\mathfrak{r} \in p))\text{”} \tag{δ}$$

or

$$\text{“}\forall\mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg e \in (\mathfrak{r} \in p)) \rightarrow \neg e \in (a \in (p|q))\text{”} \tag{ϵ}$$

By drawing on this sentence one passes from

$$\begin{aligned}&\text{“I}p \rightarrow (\text{I}q \rightarrow (e \in (d \in (p|q)) \rightarrow (\neg d = a\\ &\quad \rightarrow \forall \mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg e \in (\mathfrak{r} \in p)))))\text{”}\end{aligned} \tag{ζ}$$

to the sentence

$$\begin{aligned}&\text{“I}p \rightarrow (\text{I}q \rightarrow (e \in (d \in (p|q))\\ &\quad \rightarrow (\neg d = a \rightarrow \neg e \in (a \in (p|q)))))\text{”}\end{aligned} \tag{η}$$

from which (γ) follows by contraposition. (ζ) is obtained by rule (5) from

$$\begin{aligned}&\text{“I}p \rightarrow (\text{I}q \rightarrow (e \in (d \in (p|q))\\ &\quad \rightarrow (\neg d = a \rightarrow (b \in (a \in q) \rightarrow \neg e \in (b \in p)))))\text{”}.\end{aligned} \tag{θ}$$

This formula must be derived by means of

$$\text{“}\forall \mathfrak{r}(\mathfrak{r} \in (d \in q) \rightarrow \neg e \in (\mathfrak{r} \in p)) \rightarrow \neg e \in (d \in (p|q))\text{”} \tag{ϵ}$$

from

$$\begin{aligned}&\text{“I}p \rightarrow (e \in (b \in p) \rightarrow (\neg d = a \rightarrow (\text{I}q \rightarrow (b \in (a \in q)\\ &\quad \rightarrow (c \in (d \in q) \rightarrow \neg e \in (c \in p))))))\text{”}\end{aligned} \tag{ι}$$

in a way similar to the way in which (η) is derived from (θ), bearing in mind that “c” needs to be replaced by “$\mathfrak{r}$”.

$$\text{“I}q \rightarrow (b \in (d \in q) \rightarrow (b \in (a \in q) \rightarrow d = a\text{”} \tag{κ}$$

follows from the definition of *single-valuedness of a relation* (Def. Γ), and from this by means of (IIIc)

$$\text{“}b = c \rightarrow (\text{I}q \rightarrow (c \in (d \in q) \rightarrow (b \in (a \in q) \rightarrow d = a)))\text{”} \tag{λ}$$

If one applies (κ) to this in the form

$$\text{“}p \rightarrow (e \in (b \in p) \rightarrow (e \in (c \in p) \rightarrow b = c))\text{”}, \tag{κ}$$

then one succeeds in proving the formula (ι), with the help of which one can obtain (α).

(b) Proof of the Formula

$$\begin{aligned}&\text{“}v \in (u \in \rangle \int q) \rightarrow (u \in (v \in \rangle q) \rightarrow \neg \forall \mathfrak{q}(v \in (w \in \rangle \int \mathfrak{q})\\ &\quad \rightarrow \neg w \in (v \in \rangle \mathfrak{q})) \rightarrow (\neg \forall \mathfrak{q}(u \in (w \in \rangle \int \mathfrak{q}) \rightarrow \neg w \in (u \in \rangle \mathfrak{q}))))\text{”}\end{aligned}$$

(cf. analysis 1, sentence ε)

and end of section A. Proof of Hume's Principle

Analysis 4 (§60)
Frege must now prove (ν) of §54, namely

$$\text{“}\int(p|q) = \int(q|\int p)\text{”}. \tag{α}$$

According to the definitions of *converse of a relation* (Def. E) and *composite relation* (Def. B), this amounts to proving

$$\begin{aligned}&\text{“VR}\alpha\text{VR}\varepsilon(\alpha \in (\varepsilon \in (p|q)\\ &\quad = \text{VR}\alpha\text{VR}\varepsilon(\neg\forall\mathfrak{r}(\mathfrak{r} \in (\alpha \in \int p) \rightarrow \neg\varepsilon \in (\mathfrak{r} \in \int q)))\text{”}.\end{aligned} \tag{β}$$

To carry out this step Frege can use the formula

$$\begin{aligned}&\text{“}\forall\mathfrak{d}\forall\mathfrak{a}(f(\mathfrak{a},\mathfrak{d}) = g(\mathfrak{a},\mathfrak{d}))\\ &\quad \rightarrow \text{VR}\alpha\text{VR}\varepsilon(f(\varepsilon,\alpha)) = \text{VR}\alpha\text{VR}\varepsilon(g(\varepsilon,\alpha))\text{”}\end{aligned} \tag{γ}$$

which in turn can be proved by applying twice (Va). In order to apply (Va) here, Frege needs

$$\text{“}b \in (a \in (p|q)) = (\neg\forall\mathfrak{r}(\mathfrak{r} \in (b \in \int p) \rightarrow \neg a \in (\mathfrak{r} \in \int q)))\text{”} \tag{δ}$$

which follows by (4) from

$$\begin{aligned}&\text{“}(\neg\forall\mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg b \in (\mathfrak{r} \in \int p))\\ &\quad \rightarrow \neg\forall\mathfrak{r}(\mathfrak{r} \in (b \in \int p) \rightarrow \neg a \in (\mathfrak{r} \in \int q)))\text{”}.\end{aligned} \tag{ϵ}$$

(e) must be derived by means of (IVa). For this, Frege needs

$$\begin{aligned}&\text{“}\neg\forall\mathfrak{r}(\mathfrak{r} \in (b \in \int p) \rightarrow \neg a \in (\mathfrak{r} \in \int q)))\\ &\quad \rightarrow \neg\forall\mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg b \in (\mathfrak{r} \in p))\text{”}\end{aligned} \tag{ζ}$$

and

$$\begin{aligned}&\text{“}\neg\forall\mathfrak{r}(\mathfrak{r} \in (a \in q) \rightarrow \neg b \in (\mathfrak{r} \in p))\\ &\quad \rightarrow \neg\forall\mathfrak{r}(\mathfrak{r} \in (b \in \int p) \rightarrow \neg a \in (\mathfrak{r} \in \int q))\text{”}.\end{aligned} \tag{η}$$

Both (ζ) and (η) can be derived from

$$\text{“}\mathfrak{r} \in (a \in q) = a \in (\mathfrak{r} \in \int q)\text{”}$$

which follows from (E). Frege draws upon the sentence (α) that has been proved in this way in the proof of

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \rightarrow u \in (w \in\rangle \int p)\\ &\quad \rightarrow v \in (w \in\rangle \int (p|q))\text{”(cf. §54);}\end{aligned} \tag{ι}$$

and from this and (19)

$$\text{“}w \in (u \in\rangle p) \rightarrow (u \in (v \in\rangle q) \rightarrow w \in (v \in\rangle p|q)))\text{”}$$

he derives (ε) (see the end of construction 3, §59) (cf. §54).

Analysis 5 (§62)
In order to derive (δ) (§54)

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \rightarrow (u \in (v \in\rangle q) \rightarrow \neg\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q})\\ &\quad \rightarrow \neg w \in (v \in\rangle \mathfrak{q})) \rightarrow (\neg\forall \mathfrak{q}(u \in (w \in\rangle \int \mathfrak{q}) \rightarrow \neg w \in (u \in\rangle \mathfrak{q}))))\text{”}\end{aligned}$$

from (25) (cf. end of construction 4, §61)

$$\begin{aligned}&\text{“}v \in (u \in\rangle \int q) \rightarrow (u \in (v \in\rangle q) \rightarrow (\neg\forall \mathfrak{q}(u \in (w \in\rangle \int \mathfrak{q})\\ &\quad \rightarrow \neg w \in (u \in\rangle \mathfrak{q})) \rightarrow \neg\forall \mathfrak{q}(v \in (w \in\rangle \int \mathfrak{q}) \rightarrow \neg w \in (v \in\rangle \mathfrak{q}))))\text{”},\end{aligned}$$

Frege needs (η) (§54)

$$\text{“}u \in (v \in\rangle \int q) \rightarrow \neg u \in (v \in\rangle \int\int q)\text{”}. \tag{α}$$

According to (11) (cf. construction 2, §57),

$$\begin{aligned}&\text{“I}q \rightarrow (\forall \mathfrak{d}(\forall \mathfrak{a}(\mathfrak{d} \in (\mathfrak{a} \in q \rightarrow \neg \mathfrak{a} \in v)\\ &\quad \rightarrow \neg \mathfrak{d} \in w) \rightarrow w \in (v \in\rangle q))\text{”},\end{aligned}$$

both

$$\begin{aligned}&\text{“}u \in (v \in\rangle \int q) \rightarrow \forall \mathfrak{d}(\forall \mathfrak{a}(\mathfrak{d} \in (\mathfrak{a} \in \int\int q)\\ &\quad \rightarrow \neg \mathfrak{a} \in v) \rightarrow \neg \mathfrak{d} \in u)\text{”}\end{aligned} \tag{β}$$

and

$$\text{“}u \in (v \in\rangle q) \rightarrow \text{I}\int\int q\text{”} \tag{γ}$$

require proof.

(β) emerges from

$$\text{“}u \in (v \in\rangle q) \rightarrow (\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\int q) \rightarrow \neg \mathfrak{a} \in v) \rightarrow \neg d \in u)\text{”} \tag{δ}$$

by applying rule (5).

According to (8) (cf. construction 2, §57)

$$\text{“}u \in (v \in\rangle q) \to (\forall \mathfrak{a}(e \in (\mathfrak{a} \in q) \to \neg \mathfrak{a} \in v) \to \neg e \in u)\text{”},$$

we now have

$$\text{“}u \in (v \in\rangle q) \to (\forall \mathfrak{a}(d \in (\mathfrak{a} \in q) \to \neg \mathfrak{a} \in v) \to \neg d \in u)\text{”}. \qquad (\epsilon)$$

Thus, it remains to be proved

$$\begin{array}{l} \text{“}u \in (v \in\rangle q) \to (\forall \mathfrak{a}(d \in (\mathfrak{a} \in q) \to \neg \mathfrak{a} \in v) \to \neg d \in u)\text{”} \\ \text{“}(\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\!\!\int q) \to \neg \mathfrak{a} \in v) \to (\forall \mathfrak{a}(d \in (\mathfrak{a} \in q) \to \neg \mathfrak{a} \in v)\text{”} \end{array} \qquad (\zeta)$$

(ζ) follows by rule (5) from

$$\text{“}(\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\!\!\int q) \to \neg \mathfrak{a} \in v) \to (d \in (\mathfrak{a} \in q) \to \neg \mathfrak{a} \in v)\text{”}. \qquad (\eta)$$

If we transform (IIa) into

$$\text{“}(\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\!\!\int q) \to \neg \mathfrak{a} \in v) \to (d \in (\mathfrak{a} \in \int\!\!\int q) \to \neg \mathfrak{a} \in v)\text{”}, \qquad (\theta)$$

we see that

$$\text{“}d \in (\mathfrak{a} \in q) \to d \in (\mathfrak{a} \in \int\!\!\int q)\text{”}$$

is provable with the aid of (22); cf. construction 4, §61.

Construction 5 (§63)

$$(22) \quad a \in (d \in \int q) \to d \in (a \in \int\!\!\int q)$$

(22) :: – – – – – – – – –

$$d \in (a \in q) \to d \in (a \in \int\!\!\int q) \qquad (26)$$

(IIa) : – – – – – – – –

$$\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\!\!\int q) \to \neg \mathfrak{a} \in v) \to (d \in (\mathfrak{a} \in q) \to a\neg \in v) \qquad (\alpha)$$

$$\div$$

$$\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\!\!\int q) \to \neg \mathfrak{a} \in v) \to \forall \mathfrak{a}(d \in (\mathfrak{a} \in q) \to \neg \mathfrak{a} \in v) \qquad (\beta)$$

(8) : – – – – – – – –

$$u \in (v \in\rangle q) \to (\forall \mathfrak{a}(d \in (\mathfrak{a} \in \int\int q) \to \neg \mathfrak{a} \in v) \to \neg d \in u) \qquad (\gamma)$$

$$\div$$

$$u \in (v \in\rangle q) \to (\forall \mathfrak{d}(\forall \mathfrak{a}(\mathfrak{d} \in (\mathfrak{a} \in \int\int q \to \neg \mathfrak{a} \in v) \to \neg \mathfrak{d} \in u) \qquad (\delta)$$

(11): – – – – – – – –

$$\mathrm{I}\int\int q \to u \in (v \in\rangle q) \to u \in (v \in\rangle \in \int\int q) \qquad (27)$$

Analysis 6 (§64)
In order to carry out the last step (=construction 6) in the proof of Hume's Principle, Frege still needs to prove "$u \in (v \in\rangle q) \to \mathrm{I}_{\int\int} q$" (formula (γ) of §62). After having accomplished this, he can use (γ) for the purpose of eliminating in (27) the antecedent by way of fusing the antecedents. When this is done, Frege arrives finally at Hume's Principle.

In the first place, he proves

$$\text{“}\mathrm{I}q \to \mathrm{I}\int\int q\text{”} \qquad (\alpha)$$

from which (γ) follows together with (18) (cf. construction 3, §59)

$$\text{“}u \in (v \in\rangle q) \to \mathrm{I}q\text{”}.$$

According to (16), it is mandatory to prove

$$\begin{aligned}&\text{“}\mathrm{I}q \to (\forall \mathfrak{e}(\forall \mathfrak{d}(\mathfrak{e} \in (\mathfrak{d} \in \int\int q \to \forall \mathfrak{a}(\mathfrak{e} \in (\mathfrak{a} \in \int\int q)\\ &\quad \to \mathfrak{d} = \mathfrak{a}\neg \mathfrak{a} \in v) \to \neg \mathfrak{d} \in u))\text{”}\end{aligned} \qquad (\beta)$$

or

$$\text{“}\mathrm{I}q \to (e \in (d \in \int\int q) \to e \in (a \in \int\int q) \to d = a)\text{”} \qquad (\gamma)$$

According to (13), one now has

(γ) follows from this and the sentence

$$\text{“}e \in (a \in \int\int q) \to e \in (a \in q)\text{”} \qquad (\epsilon)$$

which follows from (23) (cf. construction 4, §61)

$$\text{“}F(a \in (r \in \int q)) \to F(r \in (a \in q))\text{”}$$

in a way similar to that in which (26) follows from (22).

Construction 6 (§65)

$$23 \quad (e \in (a \in \int\int q)) \to a \in (e \in \int q))$$

(23) : – – – – – – – –

$$e \in (a \in \int\int q)) \to e \in (a \in \int q) \tag{28}$$

(13) : – – – – – – – –

$$\mathrm{I}q \to (e \in (d \in q) \to (e \in (a \in \int\int q) \to d = a)) \tag{α}$$

(28) : – – – – – – – –

$$\mathrm{I}q \to (e \in (d \in \int\int q) \to (e \in (a \in \int\int q) \to d = a)) \tag{β}$$

$$\div$$

$$\mathrm{I}q \to \forall \mathfrak{e} \forall \mathfrak{d} (\mathfrak{e} \in (\mathfrak{d} \in \int\int q) \to \forall \mathfrak{a} (\mathfrak{e} \in (\mathfrak{a} \in \int\int q) \to \mathfrak{d} = \mathfrak{a} \tag{γ}$$

(16) : – – – – – – – –

$$\mathrm{I}q \to \mathrm{I}\int\int q \tag{29}$$

(18) : – – – – – – – –

$$u \in (v \in\rangle q) \to \mathrm{I}\int\int q \tag{30}$$

(27) : – – – – – – – –

$$u \in (v \in\rangle q) \to u \in (v \in\rangle \mathrm{I}\int\int q) \tag{31}$$

(25) : – – – – – – – –

$$\begin{aligned} u \in (v \in\rangle q) &\to (v \in (u \in\rangle \int q) \\ &\to (\neg \forall \mathfrak{q} (v \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (v \in\rangle \mathfrak{q})) \\ &\to \neg \forall \mathfrak{q} (u \in (w \in\rangle \int \mathfrak{q}) \to \neg w \in (u \in\rangle \mathfrak{q})))) \end{aligned} \tag{α}$$

(IVa) : – – – – – – – –

$$
\begin{aligned}
& u \in (v \in\rangle q) \rightarrow (v \in (u \in\rangle \textstyle\int q) \\
& \quad \rightarrow ((\neg\forall \mathfrak{q}(u \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg w \in (v \in\rangle \mathfrak{q})) \\
& \quad \rightarrow \neg\forall \mathfrak{q}(v \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg w \in (v \in\rangle \mathfrak{q}))) \\
& \quad \rightarrow (\neg\forall \mathfrak{q}(u \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg w \in (u \in\rangle \mathfrak{q}))) \\
& \quad \rightarrow (\neg\forall \mathfrak{q}(v \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg w \in (v \in\rangle \mathfrak{q})))
\end{aligned} \tag{β}
$$

(25) : – – – – – – – –

$$
\begin{aligned}
& v \in (u \in\rangle q) \rightarrow (u \in (v \in \textstyle\int q) \\
& \quad \rightarrow (\neg\forall \mathfrak{q}(u \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg w \in (u \in\rangle \mathfrak{q}))) \\
& \quad \rightarrow (\neg\forall \mathfrak{q}(v \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg w \in (v \in\rangle \mathfrak{q})))
\end{aligned} \tag{γ}
$$

$$\div$$

$$
\begin{aligned}
& v \in (u \in\rangle q) \rightarrow (u \in (v \in\rangle \textstyle\int q) \\
& \quad \rightarrow \forall \mathfrak{a}((\neg\forall \mathfrak{q}(u \in (\mathfrak{a} \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg \mathfrak{a} \in (u \in\rangle \mathfrak{q}))) \\
& \quad = \neg\forall \mathfrak{q}(v \in (w \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg \mathfrak{a} \in (v \in\rangle \mathfrak{q}))))
\end{aligned} \tag{δ}
$$

(Va) : – – – – – – – –

$$
\begin{aligned}
& v \in (u \in\rangle q) \rightarrow (u \in (v \in\rangle \textstyle\int q) \\
& \quad \rightarrow \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(u \in (\varepsilon \in \textstyle\int \mathfrak{q}) \rightarrow \neg \varepsilon \in (u \in\rangle \mathfrak{q}))) \\
& \quad = \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(v \in (\varepsilon \in \textstyle\int \mathfrak{q}) \rightarrow \neg \varepsilon \in (v \in\rangle \mathfrak{q}))))
\end{aligned} \tag{ϵ}
$$

(IIIc) : ——————

$$
\begin{aligned}
& \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(u \in (\varepsilon \in\rangle \textstyle\int \mathfrak{q}) \rightarrow \neg \varepsilon \in (u \in\rangle \mathfrak{q}))) = \mathrm{N}u \\
& \quad \rightarrow v \in (u \in\rangle \textstyle\int q) \rightarrow (u \in (v \in\rangle q) \\
& \quad \rightarrow \mathrm{N}u = \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(v \in (\varepsilon \in \textstyle\int \mathfrak{q}) \rightarrow \neg \varepsilon \in (v \in\rangle \mathfrak{q})))))
\end{aligned} \tag{ζ}
$$

(Z) :: ——————

$$
\begin{aligned}
& v \in (u \in\rangle \textstyle\int q) \rightarrow (u \in (v \in\rangle q) \\
& \quad \rightarrow \mathrm{N}u = \mathrm{VR}\varepsilon(\neg\forall \mathfrak{q}(v \in (\varepsilon \in \textstyle\int \mathfrak{q}) \rightarrow \neg \varepsilon \in (v \in\rangle \mathfrak{q}))))
\end{aligned} \tag{η}
$$

(IIIc): ————————

$$\begin{aligned} &\mathrm{VR}\varepsilon(\neg\forall\mathfrak{q}(v \in (\varepsilon \in\rangle\int\mathfrak{q}) \rightarrow \neg\varepsilon \in (v \in\rangle\mathfrak{q}))) = \mathrm{N}v \\ &\quad\rightarrow (v \in (u \in\rangle\int q) \rightarrow (u \in (v \in\rangle q) \rightarrow \mathrm{N}u = \mathrm{N}v)) \end{aligned} \tag{θ}$$

(Z) :: ————————

$$v \in (u \in\rangle\int q) \rightarrow (u \in (v \in\rangle q) \rightarrow \mathrm{N}u = \mathrm{N}v) \tag{32}$$

So much for Frege's formal proof of Hume's Principle. In what follows, I shall briefly comment on the criteria of identity incorporated in Hume's Principle and Basic Law V and their relation to one another.

Equinumerosity and Coextensiveness: Hume's Principle and Basic Law V Again

The criterion of identity supplied by Basic Law V applies to a larger class of objects than the criterion embedded in Hume's Principle. While the latter applies only to cardinal numbers, the former applies to logical objects à la Frege in general and, hence, also to the cardinals qua equivalence classes of equinumerosity and to the reals qua Relations of Relations. If we consider the identity conditions on the right-hand side of Axiom V, we see that they are more tightly woven than those on the right-hand side of Hume's Principle. Plainly, the coextensiveness of two concepts $F(x)$ and $G(x)$ implies the equinumerosity of $F(x)$ and G(x): $\forall x(F(x) \leftrightarrow G(x)) \rightarrow Eq_x(F(x), G(x))$, and by virtue of Hume's Principle also $N_xF(x) = N_xG(x)$, but the converse does not hold.

Someone might wish to argue that in the system of *Grundgesetze* equinumerosity between concepts (or their extensions) is no longer needed in its function as governing the identity conditions of cardinal numbers; this task is now assigned to and taken care of by Axiom V. Now it is undeniable that Frege badly needs the relation of equinumerosity for his definition of the cardinality operator as well as for his proof of Hume's Principle from that definition. The cardinality operator (the name of the cardinality function) refers now to a monadic *first-level* function, but the new definition is modelled upon the pattern of the old one, its famous predecessor in *Grundlagen*, §68. In part II of Frege (1893), after having completed the exposition of the concept-script in part I, Frege turns to the proofs of the basic laws of cardinal number and, as we have seen, it is Hume's Principle that he proves first. It almost goes without saying that in every proof of a fundamental law of cardinal arithmetic based on Hume's Principle—for example, in the proof of the proposition "that the relation of a cardinal number to the one immediately following it is single-valued" (Frege 1893, §66)—equinumerosity is essentially involved in its

function as supplying the identity conditions for cardinal numbers.[28] In a case like this, Hume's Principle could of course not be replaced by Axiom V. By contrast, if we only wish to establish whether the cardinal number that belongs to (the extension of) the concept *F(x)*—qua extension of the concept *equinumerous with the concept F(x)* or qua extension of the concept *equinumerous with the extension of F(x)*—is equal to the cardinal number that belongs to (the extension of) the concept *G(x)*—qua extension of the concept *equinumerous with the concept G(x)* or qua extension of the concept *equinumerous with the extension of G(x)*—we have the choice to do this by appealing either to Hume's Principle or to Axiom V or to the third-order cousin of Axiom V to which I drew attention in Section "The

[28]As Frege puts it, "the relation in which one member of the cardinal number series stands to that immediately following it" is defined in Frege (1893), §43 (Def. H) as follows:

$$\begin{aligned}\mathrm{f} := \mathrm{VR}\alpha\mathrm{VR}\varepsilon(\neg\forall\mathfrak{u}\forall\mathfrak{a}(\mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{u} \rightarrow \varepsilon = a)))\\ = \varepsilon \rightarrow (a \in \mathfrak{u} \rightarrow \neg\mathrm{N}\mathfrak{u} = \alpha))).\end{aligned}$$

From Def. H one can easily derive

$$\begin{aligned}e \in (a \in \mathrm{f}) \rightarrow (\neg\forall\mathfrak{u}\forall a(\mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{u} \rightarrow \varepsilon = a))) = \varepsilon\\ \rightarrow (a \in \mathfrak{u} \rightarrow \neg\mathrm{N}\mathfrak{u} = a)).\end{aligned}$$

What needs to be proved regarding the single-valuedness of f is therefore

$$\begin{aligned}\neg\forall\mathfrak{u}\forall a(\mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{u} \rightarrow \varepsilon = a))) = \varepsilon \rightarrow (a \in \mathfrak{u} \rightarrow \neg\mathrm{N}\mathfrak{u} = d))\\ \rightarrow \neg\forall\mathfrak{u}\forall a(\mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{u} \rightarrow \varepsilon = a))) = \varepsilon \rightarrow (a \in \mathfrak{u} \rightarrow \neg\mathrm{N}\mathfrak{u} = a).\end{aligned}$$

This is a formula which, by repeated application of contraposition and the introduction of German letters, emerges from

$$\begin{aligned}\mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{v} \rightarrow \varepsilon = c))) = \varepsilon \rightarrow (c \in v \rightarrow \mathrm{N}\mathfrak{v} = d\\ \rightarrow \mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{u} \rightarrow \varepsilon = b))) = \varepsilon \rightarrow (b \in u \rightarrow (\mathrm{N}u = a \rightarrow d = a))))).\end{aligned}$$

The latter formula can be derived from

$$\begin{aligned}c \in v \rightarrow (\mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in u) \rightarrow \varepsilon = b))) = \mathrm{N}(\mathrm{VR}\varepsilon(\neg(\varepsilon \in \mathfrak{v} \rightarrow \varepsilon = c)))\\ \rightarrow (b \in u \rightarrow \mathrm{N}u = \mathrm{N}\mathfrak{v})).\end{aligned}$$

It is at this point of his proof of the single-valuedness of f that Frege expressly invokes Hume's Principle. According to Hume's Principle, the last-mentioned derivation can be carried out by showing that there is a relation which maps the *u*-concept into the *v*-concept and whose converse maps the *v*-concept into the *u*-concept. And that there is indeed a relation which maps the $\mathrm{VR}\varepsilon(\neg(\varepsilon \in\rightarrow \varepsilon = b))$-concept into the $\mathrm{VR}\varepsilon(\neg(\varepsilon \in v \rightarrow \varepsilon = c))$-concept follows from the identity of the cardinal number belonging to the first concept with that belonging to the second, which of course must still be proved.

Proof Sketch in *Grundlagen*". In such a case, the criterion of identity of equinumerosity and that of coextensiveness work equally well.[29]

Julius Caesar and Cardinal Numbers—A Brief Comparison Between *Grundlagen* and *Grundgesetze*

In spite of all the similarity between the *Grundlagen* and the *Grundgesetze* approaches to cardinal arithmetic, it would be patently false to transfer the close connection that we find in *Grundlagen* between the emergence of the Caesar problem from Hume's Principle and its attempted solution via the explicit definition of the cardinality operator to Frege (1893). As I already said, the situation in Frege (1893) differs significantly from that in Frege (1884) as far as the Caesar problem and its alleged solution are concerned.

Firstly, the Caesar problem arising from Hume's Principle in Frege (1884) is not even mentioned in Frege (1893), neither before Frege defines explicitly the cardinality operator nor when he turns to Hume's Principle and sets out to prove it; and this is not merely an accident.

Secondly, if in *Grundgesetze* this explicit definition were intended to resolve a formal variant of the original Caesar problem from *Grundlagen*, relating to cardinal numbers, then that problem would probably have to emerge from a tentative contextual definition of the cardinality operator just as it did in *Grundlagen*. However, since in *Grundgesetze* contextual definitions are strictly prohibited, that proves to be impossible. Imagine now the period when Frege embarked on writing Frege (1893) and suppose, for the sake of argument, that he installed Hume's Principle as an axiom governing the cardinality operator as one of the primitive signs of his concept-script. In that case, he would indeed have encountered a formal version of the Caesar problem arising from the tentative contextual definition of the cardinality operator in *Grundlagen*. Yet the possibility of resolving it by putting forward the explicit definition of the cardinality operator would be unavailable. On my assumption, "$N_x\varphi(x)$" would be primitive and therefore indefinable in the relevant setting.

Thirdly, as was already observed, when in Frege (1893) Frege introduces value-ranges and value-range terms he faces a Caesar problem in a formal guise. He thinks that it is solved once the method that he proposes in §10—to determine the values of every primitive first-level function for value-ranges as well as for all other admissible arguments—is completely carried out. As we have seen, it is only Frege (1893), §53 that he introduces Hume's Principle, presenting now a new version of it

[29]Thus, following Frege's strategy in *Grundlagen*, we have the choice between "$N_xF(x) = N_xG(x) \leftrightarrow Eq_x(F(x), G(x))$" (or its definitional variant "$Ext\varphi(Eq_x(\varphi(x), F(x))) = Ext\varphi(Eq_x(\varphi(x), G(x))) \leftrightarrow Eq_x(F(x), G(x))$") and "$Ext\varphi(Eq_x(\varphi(x), F(x))) = Ext\varphi(Eq_x(\varphi(x), G(x))) \leftrightarrow \forall\varphi(Eq_x(\varphi(x), F(x)) \leftrightarrow Eq_x(\varphi(x), G(x)))$", whenever we wish to establish whether the number that belongs to $F(x)$ is equal to the number that belongs to $G(x)$.

which is otherwise equivalent to the old one in *Grundlagen*. At this stage, cardinal numbers are already defined as special value-ranges, as equivalence classes of equinumerosity. Thus, from Frege's point of view, by fixing completely the references of value-range terms and by subsequently identifying cardinal numbers with value-ranges he has succeeded in fixing uniquely the references of numerical terms standing for cardinals as well. Moreover, casting an eye over Frege's strategy of laying the logical foundations of real analysis in Frege (1903), he might have argued as follows. First, an additional axiom (besides Axiom V) governing extensions of (first-level) relations, or more generally, value-ranges of dyadic (first-level) functions (the so-called double value-ranges), is not required.[30] Second, once the reals are defined as Relations of Relations,[31] the references of the standard terms for the reals are likewise uniquely fixed. In Frege (1893), a Caesar problem simply does not arise from Hume's Principle or to put it in slightly paradoxical terms: it is already solved before it could arise. Once Hume's Principle is introduced in §53, the task is then to derive it from the explicit definition of the cardinality operator and to establish it in this way as a theorem of logic.

Cardinals and Reals by Abstraction: *Grundlagen* and *Grundgesetze*

To conclude this essay, let me cast a glance at Frege's logicist project in *Grundlagen* and in *Grundgesetze* from a another point of view. In *Grundlagen*, §104, Frege made it clear that in potential subsequent work he intended to introduce the real and complex numbers along the lines of his introduction of the cardinals, namely by starting with a tentative contextual definition of a suitable number operator in terms of an abstraction principle whose right-hand side, a second-order or higher-order equivalence relation, was framed in purely logical terms (and thus providing logical criteria of identity for the real or complex numbers) and by finally defining these numbers as equivalence classes of the relevant equivalence relation. In the case of the cardinal numbers and especially in that of the real numbers, Frege takes a different route in *Grundgesetz*e. As to the cardinals, it is true that both in *Grundlagen* and in Frege (1893) he defines them as equivalence classes of equinumerosity and it is likewise true that in the two works he treats Hume's Principle as the fulcrum of the proofs of the fundamental laws of number theory. Yet thanks to the introduction in Frege (1893) of value-ranges qua logical target objects for the definitions of numbers of all kinds, Frege no longer thought that it

[30]See Heck (1997, p. 283 f.).

[31]The term "*Relation*" is Frege's shorthand expression for "*Umfang einer Beziehung*" ("extension of a relation"). Thus, in his logic *Relationen* (in English, I use "Relations" with a capital "R") are value-ranges of two-place (first-level) functions whose value, for every pair of admissible arguments (objects), is either the True or the False. Note that in Frege (1903), Frege did not yet set up a definition of the reals. He only informs us about the way how he wants to define them.

was methodologically requisite or useful to introduce cardinal numbers exactly along the lines of the *Grundlagen* strategy.

Regarding the projected definition of the real numbers and the proof of the fundamental laws of analysis in Frege (1903) (and in a planned third *Grundgesetze* volume), the methodological approach differs at least in one essential respect from that concerning the logical foundations of cardinal arithmetic in Frege (1893). In contrast to his treatment of the cardinal numbers, Frege does not introduce any abstraction principle for the real numbers with the intention of establishing it as a truth of logic by way of proving it from a definition of a real number operator whose definiens was couched in purely logical vocabulary. Plainly, such a principle would have furnished logical criteria of identity for the reals and might have played a role in the foundations of real analysis similar to the role that Hume's Principle was designed to play in the foundations of number theory. Yet it seems that when writing *Grundgesetze* Frege did not see any requirement to introduce a special abstraction principle for the real numbers as long as he thought that Axiom V was trustworthy and could provide the appropiate cognitive access to all numbers, not only to the cardinals. Following his plan, once the real numbers qua "measurement numbers" had been defined as Relations of Relations, Axiom V would take care of the identity conditions for the reals.[32] So, I think that an abstraction principle figuring as a key theorem in real analysis and as such governing the identity conditions of the real numbers would not have been forthcoming in a third *Grundgesetze* volume, had Frege seen a chance of bringing his project of laying the logical foundations of analysis to a happy ending.

There is another speculative question that may suggest itself in this context. What would have been the prospect for Frege after 1902—when he was facing the disastrous consequences of Russell's Paradox regarding the feasibility of his logicist project—for the introduction of the real numbers via an abstraction principle? (In what follows, I refer to the hypothetical abstraction principle for the reals as "APR".) Could he have contrived and laid down such a principle without betraying his logicist credo? To answer this question tentatively, recall that in *Grundgesetze*, in the light of Frege's theory of definition for his formal language, abstraction principles are ruled out once and for all in their originally proposed function in *Grundlagen* as definitions of term-forming operators. This holds quite independently of the Julius Caesar problem which, as we have seen, affects Fregean abstraction principles across the board, regardless of whether such a principle appears in the guise of a (tentative) contextual definition, as does initially Hume's Principle in *Grundlagen*, or is clad in the garb of an axiom, as is the case with the transformation of the coextensiveness of two monadic first-level functions into an identity of value-ranges (and vice versa) in Frege (1893). As I pointed out earlier, the reason that in *Grundgesetze* any definition in terms of an abstraction principle is

[32]As regards Frege's theory of real numbers, see von Kutschera (1966), Simons (1987), Dummett (1991) and Schirn (2013, 2014a).

inadmissible, is that contextual definitions infringe upon Frege's principles of correct definitions which were still absent in *Grundlagen*.

Here, then, is my answer to the question whether in the aftermath of Russell's Paradox Frege could have introduced the real numbers via an abstraction principle APR without offending against the spirit of logicism. It seems to me that in principle he would have been in a position to do this, if he thought that he had compelling grounds for accepting APR as a primitive truth of logic. The grounds would be compelling, from his point of view, if APR met four requirements: (1) it would have to be true; (2) it would have to be self-evident, where, in Frege's view, a self-evident truth is one that does not need deductive justification; it is supposed to be incontestable and can be acknowledged directly, in a non-inferential way; (3) it would have to possess utmost generality; (4) it would have to contain real knowledge. Once APR had passed this test successfully, nothing would initially stand in the way of laying it down as a logical axiom of the theory T of real numbers. Of course, in the end the choice could be accepted as the right one only if APR qua axiom proved to be fruitful in the sense that the fundamental theorems of T could be derived from it when augmented by definitions of certain key notions of T couched in purely logical vocabulary. And not to forget, resolving the burning Caesar problem arising almost inevitably from APR would likewise have been on the agenda. Note that according to this idea, which leaves cardinal arithmetic out of account, the logical construction of analysis would proceed without any recourse to value-ranges. Thus, if successful, the reals would enjoy the distinguished status of being logical objects in their own right.[33] Moreover, the set-theoretic paradoxes would not arise at all.

---

[33]Hale (2000) pursues the aim of providing an informal axiomatic characterization of quantitative domains, on the basis of which it will be possible to introduce the real numbers by means of an appropriate abstraction principle. Following this idea, he introduces objects—*cuts*—corresponding to cut-properties by the following abstraction principle:

$$\text{Cut} : \#F = \#G \leftrightarrow \forall a(Fa \leftrightarrow Ga)$$

where F, G are any cut-properties on $\mathrm{R}^{\mathrm{N}+}$ and a ranges over $\mathrm{R}^{\mathrm{N}+}$.

Informally, a cut-property is a non–empty property whose extension is a proper subset of $\mathrm{R}^{\mathrm{N}+}$ and which is downwards closed and has no greatest instance (cf. p. 411). Like Hume's Principle and Axiom V, Cut is a second-order abstraction principle. Yet unlike the former two principles, Cut is a restricted abstraction principle: the domain for the abstraction embraces only cut-properties on a certain specified underlying domain of objects. See Hale's discussion of cut-abstraction on p. 414 ff. If I am right, then Cut would not have been an option for Frege had he thought about the prospects of saving the intended purely logical foundation of analysis in the aftermath of Russell's Paradox. First, to use it as a contextual definition is ruled out from the outset. Second, in the light of Frege's conception of primitive laws of logic, Cut can hardly claim to be such a law. Hence, it could not be laid down as a logical axiom in the theory of real numbers. Moreover, I doubt that Cut could be shown to be analytic in any plausible sense of analyticity. (See also Wright's reflections concerning a possible extension of the neo-logicist programme to analysis in Wright 1997.) It seems to me that the neo-Fregeans have pulled out almost all the stops to make neo-logicism palatable and attractive. Ingenious and thought-provoking as some of their work undoubtedly is, I do not think that the approach deserves to be called (neo-)logicism in any well-established sense of logicism.

Acknowledgments My thanks go to the editor of this volume, Sorin Costreie, for inviting me to write an essay for it and his interest in my work on Frege. Special thanks are due to Otávio Bueno for interesting discussion and his inspiring enthusiasm about this collection.

References

Boolos, G. (1987). The consistency of Frege's foundations of arithmetic. In: J. J. Thomson (Ed.), *On being and saying. Essays for Richard Cartwright* (pp. 3–20). Cambridge, MA: The MIT Press.

Boolos, G. (1997). Is Hume's Principle analytic? In R. G. Heck (Ed.), *Language, thought, and logic. Essays in honour of Michael Dummett* (pp. 245–261).

Dummett, M. (1981). *The interpretation of Frege's philosophy*. London: Duckworth.

Dummett, M. (1991). *Frege. Philosophy of mathematics*. London: Duckworth.

Ebert, P. (2008). A puzzle about ontological commitments. *Philosophia Mathematica, 16*, 209–226.

Fine, K. (1998). The limits of abstraction. In M. Schirn (Ed.), *The philosophy of mathematics today* (pp. 503–629). Oxford: Oxford University Press.

Frege, G. (1879). *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. L. Nebert, Halle. a. S.

Frege, G. (1884). *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl.* W. Koebner, Breslau.

Frege, G. (1893). *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet* (Vol. I). Jena: H. Pohle.

Frege, G. (1903). *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet* (Vol. II). Jena: H. Pohle.

Frege, G. (1967). In I. Angelelli (Ed.), *Kleine Schriften*. Hildesheim: Georg Olms.

Frege, G. (1969). In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Nachgelassene Schriften*. Hamburg: Felix Meiner.

Frege, G. (1976). In G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, & A. Veraart (Eds.), *Wissenschaftlicher Briefwechsel*. Hamburg: Felix Meiner.

Hale, B. (2000). Reals by abstraction. *Philosophia Mathematica, 8*, 100–123; reprinted in Hale and Wright (2001) (pp. 399–420).

Hale, B., & Wright, C. (2001). *The reason's proper study. Essays towards a neo-Fregean philosophy of mathematics*. Oxford: Clarendon Press.

Heck, R. G. (1995). Frege's principle. In J. Hintikka (Ed.), *From Dedekind to Gödel* (pp. 119–142). Dordrecht: Kluwer.

Heck, R. G. (1997). The Julius Caesar objection. In R. G. Heck (Ed.), *Language, thought, and logic. Essays in honour of Michael Dummett* (pp. 273–308). Oxford: Oxford University Press.

Heck, R. G. (1999). Grundgesetze der Arithmetik I §10. *Philosophia Mathematica, 7*, 258–292.

Parsons, C. (1997). Wright on abstraction and set theory. In R. G. Heck (Ed.), *Language, thought, and logic. Essays in honour of Michael Dummett* (pp. 263–272). Oxford: Oxford University Press.

Russell, B. (1903). *The principles of mathematics* (2nd ed.). New York: Cambridge University Press.

Schirn, M. (2003). Fregean abstraction, referential indeterminacy and the logical foundations of arithmetic. *Erkenntnis, 59*, 203–232.

Schirn, M. (2006). Hume's principle and Axiom V reconsidered: Critical reflections on Frege and his interpreters. *Synthese, 148*, 171–227.

Schirn, M. (2013). Frege's approach to the foundations of analysis (1874–1903). *History and Philosophy of Logic, 34*, 266–292.

Schirn, M. (2014a). Frege's theory of real numbers in consideration of the theories of Cantor, Russell and others. In G. Link (Ed.), *Formalism and beyond. On the nature of mathematical discourse* (pp. 25–95). Boston: Walter de Gruyter.
Schirn, M. (2014b). Frege's logicism and the neo-Fregean project. *Axiomathes, 24*, 207–243.
Schirn, M. (2016a). Second-order abstraction before and after Russell's paradox. In P. Ebert & M. Rossberg (Eds.), *Essays on Frege's basic laws of arithmetic*. Oxford: Oxford University Press (forthcoming).
Schirn, M. (2016b). *Funktion, Gegenstand, Bedeutung. Freges Philosophie und Logik im Kontext*, mentis, Münster (forthcoming).
Schirn, M. (2016c). *The semantics of value-range names and Frege's proof of referentiality* (forthcoming).
Schirn, M. (2016d), Review of Gottlob Frege, *Basic Laws of Arithmetic. Derived Using Concept-Script*. Volumes I & II, P. Ebert & M. Rossberg (Trans. and Eds.). Oxford: Oxford University Press, 2013, *The Philosophical Quarterly 66*.
Simons, P. (1987). Frege's theory of real numbers. *History and Philosophy of Logic, 8*, 25–44.
von Kutschera, F. (1966). Frege's Begründung der Analysis. *Archiv für mathematische Logik und Grundlagenforschung 9*, 102–111; reprinted in Schirn (Ed.) *Studien zu Frege — Studies on Frege vol I.: Logik und Philosophie der Mathematik — Logic and Philosophy of Mathematics*, Stuttgart-Bad Cannstatt 1976 (pp. 301–312).
Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.
Wright, C. (1997). On the philosophical significance of Frege's theorem. In R. G. Heck (Ed.), *Language, thought, and logic. Essays in honour of Michael Dummett* (pp. 201–244). Oxford: Oxford University Press.
Wright, C. (1999). Is Hume's principle analytic? *Notre Dame Journal of Formal Logic, 40*, 6–30.

Author Biography

Matthias Schirn is a member of the Munich Center for Mathematical Philosophy. His research interests are in the philosophy of logic, mathematics and language, intensional semantics, epistemology and the philosophical and logical theories of Aristotle, Kant, Frege, Russell, Witttgenstein, and Hilbert. He has published several books and numerous articles in international journals and has held visiting positions at many universities including Oxford, Cambridge, Santiago de Compostela, Harvard, Berkeley, Michigan State, Minnesota, Kyoto, Buenos Aires, Lima, Mexico City, Costa Rica, Puerto Rico, Rio de Janeiro, São Paulo, Campinas, Fortaleza, João Pessoa.

Part II
Russell

A Study in Deflated Acquaintance Knowledge: Sense-Datum Theory and Perceptual Constancy

Derek H. Brown

Introduction

The heart of sense-datum theory is sense-data, a class of perceived objects that intervene between the agent and the mind-independent world in perception, rendering her knowledge of the latter indirect and acquired at least in part by virtue of sense-data representing it or being used to represent it. This is the *indirect realism* (IR) of sense-datum theory. An equally influential but in my view less essential tenet of sense-datum theory is a form of *epistemic foundationalism* (EF), the thesis that through perception we can acquire uninferred and certain justification for propositions sometimes labeled 'basic propositions'. In this case those basic propositions are about sense-data. Such a foundationalism can of course be separated from sense-datum theory,[1] but there is no doubt that many sense-datum theorists accepted and utilized sense-data in part to help bolster their commitment to it.[2]

When perceived, sense-data exist "quite as truly as anything [but] their existence and nature are to some extent dependent upon the subject" (Russell 1913, 79). This dependence of sense-data on the subject is consistent with sense-data being *mental* objects, in which case their existence depends only on the subject being in the appropriate state, or with them being *relational* objects, in which case their existence depends on both the subject and the mind-independent world being in the

[1]In recent literature see for example Fumerton (2005, 2009) and Poston (2007).

[2]Very roughly, the path to EF might proceed as follows: (a) something perceptually appears *F* to perceiver *A* iff that thing *is F*; (b) if something perceptually appears *F* to *A* then *A* can contemplate the basic proposition $\langle x \text{ is } F \rangle$; (c) given (a), (b) and some stipulations about what constitutes knowledge, if something perceptually appears *F* to *A* then *A* can know that *x* is *F*. For example Russell commits to a strong EF in *Problems of Philosophy*, and Robinson's "Phenomenal Principle" (1994, 32) can also be used to ground EF. See §§1 and 3 for discussion.

D.H. Brown (✉)
Brandon University, Brandon, Canada
e-mail: derek.brown.01@gmail.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_5

appropriate state. I will not adjudicate between mental and relational conceptions of sense-data, suffice it to say that both conceptions afford a reasonable orientation for understanding IR and EF.

In the middle of the twentieth century both tenets of sense-datum theory were heavily bombarded with criticism. IR was for example censured for postulating the existence of these intervening mind-dependent sense-data when arguments for them (e.g., drawn from the relativity of perception, perceptual illusions, and hallucinations) were scrutinized and claimed to be problematic.[3] Contrary to the sense-datum theorists' assertions, the fact that the table looks trapezoidal but is not intrinsically so offers no more reason to postulate the existence of an intervening trapezoidal sense-datum than does the fact that this electrician looks like a spy give me reason to postulate that what I see is an actual spy "between" me and this man. EF was for example criticized by pointing out that just because something is noninferentially "given" to me in perception it does not follow that I thereby have infallible knowledge of any proposition about that thing. The fact that a 38-speckled hen (or sense-datum) is presented to me in experience does not entail that I thereby have infallible knowledge of the proposition ⟨that hen (or sense-datum) has 38 speckles⟩ (see, e.g., Chisholm 1942).[4]

There today exist many responses to these objections to IR and EF, and many developed views that are opposed to both. I will not rehearse them, but will mention in passing what is helpful for our discussion. My aim is to explicate and defend a conception of IR and EF that is immune to the threats from spotted hens and from Smith's (2002) fascinating recent objection to sense-datum theory. Central to that conception is the substantive deflation of EF. Most generally, a sense-datum framework that does not rest on a robust EF has the advantage of not being immediately vulnerable to critiques of EF. The point is not so much to mandate the deflated EF for sense-datum theory, as it is to isolate a conceptual location from which sense-datum theorists and adherents of IR more generally can examine distinct perceptual phenomena to assess whether or not they *should* endorse EF for such phenomena. That is, an IR with a deflated EF contains the space to explore the extent to which EF should be adhered to, instead of the requirement to adhere to EF as much as possible.

---

[3]Perceptual relativity refers to the ways things perceptually seem to change as a host of perceptual variables change and perceptual illusion to our misperceptions of things. Hallucinations are roughly experiences of things (e.g., a dragon) that do not in fact exist in the part of the objective world being purportedly perceived.

An argument from perceptual relativity is used for example by Russell in the opening pages of *Problems of Philosophy*, by Robinson (1994) in defense of his sense-datum theory, and by Cohen (2009) to motivate his relationalist theory of colour. See Macpherson and Platchias (2013) for recent work on the difficult topic of hallucination, and the introduction to that work for an informative overview.

Criticisms and discussions of the arguments for IR can be found for example in Austin (1962), Burnyeat (1979/1980), Smith (2002), Gupta (2006) and elsewhere.

[4]More recently, Chalmers regards the speckled hen challenge as one of the core problems for what he calls "projectivism", which is fairly close to our sense-datum theory (2006, 82). See also §1.

My central case study will be shape and size constancies. I take these to pose a worthwhile challenge to sense-datum theory because of a series of observations found in Smith (2002).[5] *Perceptual constancies* roughly consist of the stability of a perceived property of a thing (e.g., its size or colour) across changes in relevant perceptual variables (e.g., distance from perceiver or conditions of illumination). They are well-known in perceptual psychology and philosophy, and beyond size and colour apply to properties such as shape and location.[6] As we will see, one way the sense-datum theorist can overcome the challenge of shape and size constancies is by deflating her EF and accepting that perceptual experience involves not merely acquaintance but also description/representation (§3). This is not the only way, but it is a way that illustrates the flexibility the deflated EF affords the sense-datum theorist. Fortunately, the needed relationship between acquaintance and representation can be drawn from some facts about ambiguity in perception (§2). The general framework I defend is centered on a particular way of carving the distinction between knowledge by acquaintance and knowledge by description (§1). It is here that IR and EF meet and from here that the constancy challenge can be charitably appreciated and overcome.

Before proceeding three remarks are in order. The extent to which I will temper EF is significant and arguably at odds with central commitments of both Russell (circa 1912) and contemporary EFs (see references in note 1). Due to space I will simply ignore these matters, and more broadly not engage in debate about whether the view defended here deserves to be called a 'sense-datum view' or an 'epistemic foundationalism'. I am less interested in defending (e.g.,) Russell or sense-datum theory *proper* than in using the broad sense-datum framework to address contemporary problems in perception. Second, deflating EF doesn't only create a more flexible IR, it also negatively impacts some simple arguments for IR (e.g., some familiar formulations of the Argument from Illusion). Individual theorists will have to weigh this 'cost' against the purported benefits of what follows. I will briefly remark on the matter in §3, but must generally leave these matters for other contributions. Third, in my view the challenge from perceptual constancies that will be posed for sense-datum theory can be equally posed for a host of presentationalist views of perception, including direct or naïve realist views and figurative

[5]I follow Siegel (2006) in interpreting Smith as utilizing constancies to generate a challenge for sense-datum theory. Smith (2006) claims that this is a misinterpretation, that he was instead intending to explain why a theory utilizing a certain conception of perceptual sensations—a conception not utilized by traditional sense-datum theory—cannot explain constancies by reference to those sensations. I hope it is obvious that regardless of Smith's intentions his remarks (quoted in §3 of this work) can be used to generate a challenge for sense-datum theory, one that I (and Siegel) believe is worth addressing.

[6]In addition to Smith (2002) recent writings in which colour constancy plays a significant role include Burge (2010), Chirimuuta (2008), Gert (2010), Hilbert (2005), Jagnow (2010), Kalderon (2008), Matthen (2010), Maund (2012). I discuss Smith because of his focus on sense-datum theory.

projectivist views.[7] Unfortunately space prevents me from doing more than highlighting the many connections between what follows and these other presentationalist views.

Acquaintance, IR and EF

Both the IR and EF tenets are often articulated and defended by reference to the idea of *acquaintance*. In its most general (and vague) form the idea is that when one comes into "direct epistemic contact" with something one is thereby acquainted with it. This contact is thought to give the agent a kind of basic epistemic access to the thing; with this contact she thereby comes to have some kind of knowledge—knowledge by acquaintance—of the thing. Understood in this way the idea of acquaintance naturally plays into both tenets of sense-datum theory. With respect to IR, because sense-data are mediators between perceivers and the mind-independent world they are the things with which we have direct perceptual contact or acquaintance, and the mind-independent world contains things with which we can and do have indirect knowledge, knowledge by description. With respect to EF, because we are acquainted with sense-data there are basic propositions about them of which we can and do possess a form of uninferred and certain knowledge. This is in contrast to the inferred and fallible knowledge by description that we possess of propositions about the mind-independent world.[8]

It is important to notice, however, that the notion of acquaintance is here being put to two very different uses. By reference to IR acquaintance constitutes a direct or immediate epistemic access that does not rest on one's epistemic access to other *things*,[9] and is contrasted with a form of knowledge that emerges as a result of one's epistemic access to some other thing. Compare for example the difference between saying "I am acquainted with Chloe" and "I've heard of Chloe". The former typically implies that I have met Chloe, that I have somehow come into contact with her. By contrast the latter implies that I know of Chloe through a less direct means, perhaps through the testimony of someone who has met her. In the latter case one's knowledge of Chloe emerges by virtue of the contact one has had with objects that

[7]Here is a rough explication. Direct realist presentationalists define perceptual experience by reference to objective things presented to perceivers (e.g., Campbell 2002; Brewer 2011). Sense-datum theorists are a kind of literal projectivist, defining perceptual experience by reference to presented sense-data and the features they instantiate. Figurative projectivists define perceptual experience by reference to presented features that are not instantiated by any presented object (i.e., they are instantiated neither by a subjective entity like sense-data nor by presented objective things). I take both forms of projectivism to be kinds of IR.

[8]It is typically additionally allowed that there are nonbasic propositions about sense-data and that at least some of these can be known by suitably endowed perceivers.

[9]I am here using 'thing' to range over objects, properties and facts.

are distinct from her; in the former case one's knowledge of Chloe emerges by virtue of the contact one has had with her.[10]

By reference to EF acquaintance yields a form of uninferred epistemically basic (and hence purportedly infallible) knowledge that is contrasted with an inferred form that involves reasoning (and hence is purportedly fallible). Consider for simplicity the difference between saying "I see blue" and "I think that thing is blue". The former statement, through a variety of qualifications, is by hypothesis directly formulable in reaction to something "given" to the agent in experience, as opposed to formulable only through a chain of reasoning that begins with or at least utilizes that experience. Hence the statement is something she cannot be wrong about. By contrast "I think that thing is blue" is by hypothesis formed through a chain of reasoning that begins with or at least utilizes some experience. This statement may but need not be correct. The relevant kind of reasoning is arguably that found at the personal as opposed to sub-personal level.[11]

Thus while both the IR and EF acquaintance notions concern "directness" and are contrasted with "mediation", the two are quite distinct. There is an ontological basis to the IR acquaintance notion that is not intrinsic to the more purely epistemic basis of the EF one: the former excludes mediation by other *things* and the latter excludes mediation by *reasoning* or *inference*. I will call 'acquaintance' as it is used by the indirect realist *Contact acquaintance*, and 'acquaintance' as it is used by the foundationalist *Noninferential acquaintance*.

Contact acquaintance is more fundamental than Noninferential acquaintance in that one can endorse the former while rejecting the latter, but the reverse is difficult to defend. Thus one can endorse the idea that there is a difference between knowing Chloe by virtue of having come into contact (being acquainted) with her and knowing her by virtue of what others have told you about her, without accepting the ideas that one possesses knowledge of her arrived at without inference or that one has infallible justification of any propositions about her (more on this shortly). On the other hand if one has infallible justification of some proposition about her it is easy to *say* that one could have arrived at it without coming into contact with her but (a) difficult to explain how this might be the case, and (b) why if it is the case the resulting knowledge should be associated with the acquaintance idea. Regarding (a), if one's knowledge of Chloe is purely through the testimony of others then maintaining that some component of that knowledge is infallible requires defending the infallibility of some kind of testimony, a prospect some may wish to embark upon but which would without question be an uphill struggle. Regarding (b), someone could for example argue that he possesses infallible justification for basic

[10]Jackson (1977) and Huemer (2001) offer two distinct means of further explicating the distinction between direct and indirect perception. These details would take us too far afield.

[11]As with the distinction between direct and indirect perception, when one delves more deeply into the distinction between basic and inferred statements or propositions various difficulties emerge. I must leave these to another time.

propositions about natural numbers and has not thereby come into "contact" with them, but instead has this knowledge through an indirect means such as the expressive powers of language. However, in making this claim it becomes immediately unclear why such knowledge would ever be deemed knowledge by acquaintance.

How well does our discussion fit with the idea of acquaintance as engineered by one of its most famous adherents, Bertrand Russell? It is well known that for Russell acquaintance knowledge is a dual-relation between epistemic agent and thing, and that descriptive knowledge is a greater-than-dual-relation that involves the agent making a judgement about that thing. Questions about truth and falsity can and can only be applied to the latter, since judgements can be correct or not; and the dual-relation he intends for acquaintance involves no judgement, hence can be neither correct nor incorrect. In this sense acquaintance knowledge doesn't *yield* a form of positive knowledge that is easy to explicate. However, in addition to this Russell famously asserted that acquaintance does yield robust knowledge, as is undeniable from the oft-quoted passage:

> The particular shade of colour that I am seeing may have many things said about it – I may say that it is brown, that it is rather dark, and so on. But such statements, though they make me know truths *about* the colour, do not make me know the colour itself any better than I did before: so far as concerns knowledge of the colour itself, as opposed to knowledge of truths about it, I know the colour perfectly and completely when I see it, and no further knowledge of it itself is even theoretically possible. (2007, 31–2)

The knowledge acquired through seeing this colour is knowledge by acquaintance. It does not contain knowledge of any truths of the thing, for that would constitute knowledge by description of it. But it does contain "perfect" and "complete" knowledge of the thing, what we might call knowledge of the thing's essence. So knowledge by acquaintance arises by coming into epistemic contact (e.g., seeing) the thing, and consists of basic and infallible (i.e., perfect and complete) knowledge of it. The roots of both Contact acquaintance and Noninferential acquaintance are here, with a fundamental qualification: whereas today EF adherents typically speak of *propositions* about a thing that are infallibly justified by being acquainted with it, Russell explicitly excludes propositional knowledge from knowledge by acquaintance. Given that propositional knowledge of a thing constitutes knowledge of truths about it, and knowledge by acquaintance is contrasted with the latter, knowledge by acquaintance must also be contrasted with propositional knowledge.

But this idea of a nonpropositional yet perfect and complete knowledge of a thing is itself somewhat problematic. If I know a thing's essence, is that knowledge not constituted by knowledge of some truth about the thing? When I am acquainted with that colour, what perfectly and completely do I know? It seems that any response will invoke some instance of propositional knowledge. Russell may reply that to demand an answer beyond "I perfectly and completely know *it*" is to prejudge the issue in favour of a propositional form of knowledge of acquaintance,

which he is at pains to avoid. In my view what the objector is looking for is further explication of what knowledge by acquaintance consists of beyond 'knowledge of the thing itself', and there are options that can be pursued: perhaps knowledge by acquaintance is not merely nonpropositional but additionally functional,[12] or perhaps it yields a discriminatory capacity (e.g., yields the ability to discriminate this thing from all others) and hence is a form of knowing *how* as opposed to the propositional knowing *that*. My aim is not to defend these or any other such means of developing the nonpropositional character of Russell's acquaintance knowledge. It is instead to stave off rejections of a broadly Russellian acquaintance knowledge on the grounds that it is nonpropositional.

What of Russell's further idea, that knowledge by acquaintance is not merely nonpropositional but perfect and complete? For my part I believe he is here on weaker ground.[13] It is not clear to me that acquaintance ever yields perfect and complete knowledge of a thing, or knowledge of a thing's essence. The EF advocate may wish to rise to his defense, to either defend perfect and complete knowledge through acquaintance, or defend infallible propositional knowledge through acquaintance of some imperfect and incomplete kind.[14] I will not venture to do so for we can admit that when I see this colour "I now know it", without committing ourselves to "I know it perfectly and completely", *and* without completely abandoning the infallibility thread of acquaintance.

What acquaintance requires, minimally, is that the thing with which the agent is acquainted exists, and that the agent has a basic kind of epistemic access to it. Regarding the former, it is simply unclear to me how one could come into epistemic contact with a thing—in the most straightforward sense in which that idea is inherent to acquaintance—without that thing existing. Thus, if that access is to noninferentially involve some epistemic state, as the EF advocate maintains, I propose that it be taken to be noninferential awareness *of the existing thing*. When I meet someone (e.g., Chloe) I thereby, without inference through things or judgement, have awareness of this existing person. Indeed Russell identifies as an "essential point" that knowledge by acquaintance "gives us our data as to what exists," and that when acquaintance occurs "the question whether there is such an object cannot arise" (1913, 80, 76, respectively). If we wish to then preserve the

[12]Gupta (2006) argues that what is epistemically given in experience should be taken to have the logical form of a function as opposed to that of a proposition. However, he is not attempting explicate knowledge by acquaintance or sense-datum theory.

[13]One could argue that Russell was not wedded to the 'perfection' and 'completeness' of acquaintance knowledge. In responding to Dawes Hick's' (1912) review of *Problems of Philosophy* Russell identified *Problems* as "a popular book where technicalities have to be avoided" and proceeded to "state as precisely as possible" his view (1913, 76). In what follows he makes no mention of these notions, and instead emphasizes the existential characterization of acquaintance I am about to introduce.

[14]Regarding the latter see again the references in Footnote 1.

infallible element endorsed by Russell and by EF advocates, it would minimally involve the assertion that one cannot be wrong about being so acquainted, that whenever one has Noninferential acquaintance with something one is thereby not only noninferentially *aware* of this existing thing, one *knows* of it.[15] Whether or not this knowledge is propositional is then open to dispute in the manner previously mentioned. For example on the propositional reading it might consist of knowledge of the proposition ⟨that thing exists⟩, and on the nonpropositional it might consist of knowledge *of the existing thing*, something which may be taken as primitive or as having a logically functional or behaviourally dispositional (etc.) form.

In a given instance this basic existential knowledge of a thing can but need not afford (using additional epistemic resources) nonbasic propositional knowledge about it and about other things. It might for example be used to infer knowledge of the proposition ⟨that thing is brown⟩. When it does so the involved acquaintance knowledge is the basis for or grounds the involved descriptive knowledge. In making this claim there is no requirement that in our perceptions acquaintance knowledge occurs first, or that it ever occurs independently of descriptive knowledge.[16] The claim is instead that these are irreducibly distinct forms of knowledge: acquaintance knowledge provides the link to perceived things which descriptive knowledge exploits. As is well known, Russell held the rather strong thesis that knowledge by acquaintance grounds *all* descriptive knowledge, making the former the foundation of his general epistemology.[17] While I have mild sympathy with this stronger claim, I will not presuppose it in what follows. I wish instead to hold that acquaintance grounds all *perceptual* descriptive knowledge, which I take to include knowledge acquired by the senses (e.g., vision, hearing, etc.) and to exclude knowledge acquired by pure thought or reflection, a qualification that ultimately requires further discussion.

Stated more formally, given my focus on perceptual knowledge I will identify and presume a *minimal acquaintance doctrine of perception* (hereafter the Doctrine). According to it a perceiver being acquainted with an object of perception consists of: a perceiver and a thing that exist (e.g., do not merely subsist) and the holding between them of the fundamental relation of acquaintance. The acquaintance relation itself consists of epistemic contact with the thing that (a) does not involve going through some other thing with which one is in epistemic contact (the Contact element), (b) does not arise through personal reasoning or inference (Noninferential element), and (c) involves the perceiver knowing of that existing

[15]One must further consider whether one is acquainted with something if and only if one is perceptually aware of it, or whether the conditional should only go in one of the two directions. I am content with the biconditional in place, but will not discuss the matter here.

[16]"[I]t would be rash to assume that human beings ever, in fact, have acquaintance with things without at the same time knowing some truth about them" (Russell 2007, 31).

[17]Recall Russell's fundamental principle regarding descriptive knowledge: "Every proposition which we can understand must be composed wholly of constituents with which we are acquainted" (2007, 40).

thing (the Infallible element).[18] This knowledge may be further construed propositionally or not.

Acquaintance is meant to provide for a perceiver an epistemic ground or anchor to a thing, a referential connection that she can exploit to acquire knowledge of various truths (i.e., knowledge by description) about that thing. The epistemic purpose of acquaintance knowledge is not itself to "tell" the agent much about what she is perceiving, it is to give her the access to what she is perceiving that is needed to formulate, through judgement, more robust and useful knowledge of perceived things. Given a secure link to the existing intentional objects of one's perceptual state, one can strive to learn about those existing things via the descriptive tools afforded by one's capacity for judgement.

This Doctrine captures a great deal of the intent of Russell's doctrine, explains the centrality of the acquaintance notion to IR and EF (and vice versa), and avoids countenancing the troublesome ideas of perfect and complete knowledge. The Doctrine is also useful because it is opposed to what is arguably the leading thought behind two of the most influential alternatives to acquaintance views offered in the twentieth century: adverbialism and intentionalism.[19] The rough idea behind these alternatives is that perceptual awareness need not involve the agent being in perceptual contact with an existing intentional object of perception, but instead need only involve the agent perceiving in a certain way (adverbialism) or being in the kind of state that represents such an object (intentionalism). These alternatives are typically initially established in domains like (non-perceptual) thought, where one can arguably think about things, such as vampires, that do not exist. A reasonable explanation of this capacity holds that this is achieved by thinking in a certain vampirely way (adverbialism) or by being in a state that represents vampires (intentionalism), thus bolstering the idea of cognitive states that do not refer. This strategy applies straightforwardly to other propositional attitudes such as desires, fears, et cetera. The relevant thought is that we should extend this strategy to perception, so that being perceptually aware does not entail the existence of an

[18]I take it as straightforward that this acquaintance doctrine satisfies a robust form of perceptual presence, and assume that "relationalists" (as, e.g., discussed in Crane 2006), and naïve and direct realists (Campbell 2002; Brewer 2011) are for our purposes individuals who see something like this minimal acquaintance doctrine (save perhaps the Infallible element) as definitive of perceptual states. Relationalists of course come in both direct and indirect realist strains—I will be focused on the latter. Acquaintance is also often taken to be a form of nonconceptual perceptual awareness. With a few exceptions to follow, I wish to remain mute on this issue. I see it working in the background in various places, but would have to greatly lengthen the work to bring them all out. [E.g., relationalists like McDowell (1994) and Brewer (2011) take the relations to the world afforded by perception to be thoroughly conceptual.] Note that Peacocke's (1983) acquaintance doctrine is distinct from mine. I regrettably do not have the space for comparison.

[19]The classic statement of adverbialism is Chisholm (1957). Four of many defenses of intentionalism are Harman (1990), Dretske (1995), Tye (2000) and Byrne (2001). One can argue (as Hilbert 2004 does) that the contemporary root of the intentionalist movement is Armstrong (1961). A possible third alternative is qualia realism (see, e.g., Block 2003; Stoljar 2004), though for example Stoljar sees this as a variant of adverbialism. See Brown (2010 and forthcoming) for a discussion of how qualia realism fits into the landscape.

intentional object of perception. This opens the door for perceptual states whose objects do not exist (hallucinations), and states some of whose "contents" are not satisfied by what is perceived (illusions). Since perception generally need not involve contact with an existing thing, let alone basic infallible knowledge of existing things, these approaches involve the denial of our Doctrine.

There is a broad sense in which I am not presently concerned with adjudicating between these alternatives. I have begun to do so elsewhere (Brown 2010, 2012, forthcoming). My present aim is to explicate and defend a plausible doctrine of acquaintance that is usable by any approach to perception that holds perception to be fundamentally relational (see Footnote 18), be such a theory of the sense-datum variety or not. My claim is that the Doctrine just articulated suffices for this purpose, and in what follows I will explain why it is needed to undercut a powerful recent objection to sense-datum theory.

Before doing so it is worth emphasizing that on this conception of acquaintance there simply are no speckled hen worries. In being acquainted with a 38-speckled hen I infallibly know of this existing thing, but not of the truth that it has 38 speckles. The hen problem arises only for acquaintance views that are not merely thoroughly propositional, but that take acquaintance to yield propositional knowledge beyond knowledge of ⟨that exists⟩, and to encompass robust knowledge of the thing's nature (e.g., knowledge of ⟨that thing has 38 speckles⟩). Regrettably, Russell held such a view,[20] as does (Fumerton 2005, 2009). This tight link between being acquainted with complex things or facts and our capacity to formulate various true "basic" propositions about them may provide a comfortable foundation for one's epistemology, but its tenability is doubtful for precisely the reasons the hen problem was put forth. It strikes me as poorly justified in Russell's case (and in Fumerton's case justified by reasons we will not delve into), and as unnecessary to the most interesting aspects of Russell's project.

Furthermore, in case it isn't obvious, commitment to such a link does not logically follow from knowledge by acquaintance *proper*, even in Russell's sense of 'acquaintance'. It requires adding an additional commitment to one's proposition forming/descriptive capacities being infallibly connected to basic aspects of what one is acquainted with. Hence, speckled hen problems target not knowledge by acquaintance but epistemologies with this additional commitment.[21] For this reason

[20]For example, in *Problems of Philosophy*, when discussing our capacity to be acquainted with complex facts Russell asserts: "In all cases where we know by acquaintance a complex fact consisting of certain terms in a certain relation, we say that the truth that these terms are so related has the first or absolute kind of self-evidence, and in these cases the judgement that the terms are so related must be true. Thus this sort of self-evidence is an absolute guarantee of truth" (2007, 99).

[21]Dretske's (1993) solution to the speckled hen problem, which emerges from his distinction between object seeing and fact seeing, is very similar to the present one, although he does not identify its connection to Russell's view. Ayer's (1940) solution (recently endorsed by Tye (2009, 2010), minus the commitment to sense-datum theory) holds that the number of apparent speckles is indeterminate. While I can envisage scenarios in which this might be useful (see, e.g., Nanay 2009) it is an odd and to my mind unsatisfying way of dealing with the problem. Unfortunately, a thorough discussion of these matters would take us too far afield.

alone there is value in exploring an IR and a conception of acquaintance knowledge that isn't wedded to these additional epistemic commitments.

We now embark on a brief foray into perceptual ambiguity, for familiarity with it is essential to appreciating the significance of the objection to sense-datum theory that is our focus.

Perceptual Ambiguity[22]

In a perceptual circumstance I take the *stimulus* to mean the object or scene as it is currently presenting itself toward the location of the perceiving agent, for simplicity the *object* and its *intrinsic features* plus the *presentational features* it occurrently offers toward the location of the agent.[23] Consider a scenario in which a perceiver finds herself in front of a wire cube oriented with the front face slightly pitched up and to the right [UR oriented], as the Necker Cube is sometimes drawn. The cube and this particular way it is presenting itself toward our agent's location is the stimulus S_1 and it contains the object O_1, the set of its intrinsic features I_1, and the set of its presentational features P_1. This set of presentational features P_1 is objectively ambiguous in the sense that various objectively different objects could present themselves toward the agent's location in a way that is perceptually indistinguishable from P_1. An obvious alternative would be a wire cube with the front face pitched down and to the left [DL oriented]. Other alternatives include a roughly two-dimensional wire figure [2D Figure] whose shape traces a flat drawing of the Necker Cube; a Stretched Cube, that is a figure with square front and back but horizontally elongated (roughly) rectangular sides and either UR or DL oriented; and so on. Each alternative constitutes a distinct stimulus S_x but each stimulus contains some set of presentational features P_x that is perceptually indistinguishable from P_1. That is, each S_x marks a candidate disambiguation of the ambiguous P_1. Call this *stimulus* ambiguity.

Perceptual ambiguity requires that the agent *see* or *perceive* a stimulus S_x as ambiguous, by which I mean that the agent sees S_x at one moment in one "way" and at another moment in a different "way". In keeping with our example, our agent sees S_1 one moment as being one thing, say a UR oriented cube, and at the next as being some other thing, say a DL oriented cube. By hypothesis this requires S_1's presentational features P_1 to be ambiguous, so that the agent can disambiguate P_1 in

[22]This section draws heavily from Brown (2012).

[23]I thus do not mean by 'stimulus' or 'given' the pattern of light reaching the eye (i.e., the retinal or proximal image; the 'sensory core' of Hatfield and Epstein 1979), although the term is sometimes used in this way. A stimulus in my sense is distal (not proximal), consisting of the objects and properties (perhaps also facts) being perceived along with the ways those entities are presenting themselves to the agent at the time of her perception. Compare with Schellenberg's (2008) 'situation-dependent properties' and the objects possessing them. Her view is discussed in Brown (2012). Also compare presentational features with Hopkins' (1998) 'outline shape'.

more than one way, that is, so that she can see S_1 as at one moment (say) S_1 and at another moment as a distinct S_2.[24]

She can of course also see S_1 as being a 2D Figure, as a UR or DL oriented Stretched Cube, et cetera. However, some of these disambiguations are perceived more readily than others. I would venture to say that seeing S_1 as a UR or DL oriented cube is easiest, and that it is roughly equally easy to see S_1 as being either of these ways. By contrast seeing S_1 as a 2D Figure is somewhat more difficult, and as a Stretched Cube (of either orientation) more difficult still. There may furthermore be other disambiguations of S_1 that the agent cannot see it as (think for example of Moretti's Blocks).

We thus have two distinct dimensions to an account of perceptual ambiguity, one consisting of the candidate disambiguations of the presentational features of the objective stimulus (the *disambiguation dimension*), and the other of the extent to which the agent can see the stimulus in accordance with each of these disambiguations (the *seeing-as dimension*). I suspect but will not argue in detail for the claim that the seeing-as dimension can involve some level of cognitive penetration: the age-relative reactions to the Dolphin illusion give decent evidence for this, as does the general fact that with practice/education it can become easier to see an ambiguous stimulus in accordance with various disambiguations. Nonetheless much work is done independently of higher-level cognitive penetrations. The fact that seeing the stimulus in our example as a UR or DL oriented cube are easiest, and roughly equally easy, suggests that our subcognitive systems have honed in on these disambiguations and judged them to be the most probably correct ones (and equally probably correct ones). I will generally say that the set of disambiguations an agent most easily sees an ambiguous stimulus as is the set of disambiguations her perceptual system judges (be it subcognitively or through both cognitive and subcognitive mechanisms) to be the most probable disambiguations of the stimulus. Disambiguations that it is harder to see the stimulus as are thus judged to be less probable, and so on.[25]

---

[24]Note that there is no requirement that in every case the subject be aware that this perceptual difference involves "flipping" between two ways of seeing a the same stimulus (as opposed to seeing different stimuli), though in many cases humans would be aware of this. There is also no requirement that the subject see the presentational features before her "on their own", that is, as a distinct set of features. Indeed I think it is unlikely that this ever occurs. Instead, I suspect that we are able to infer a set of presentational features from perceptual experience by abstraction, and only then have distinct awareness of them. I regret being unable to dwell on these matters.

[25]These judgements are likely informed by evolutionary pressures, life learning and perhaps other elements of the perceptual scene. They are thus to some degree contingent. Stimulus ambiguity of some sort is the norm in most perceptual research, be it of the sort described in the text or of the sort that identifies the stimulus with the retinal image (or a suitable analogue for other senses). One common response is to isolate operational constraints that are used or could be used by our vision system to cut down on the possible disambiguations. With respect to shape perception familiar constraints include *objects are rigid*, *objects persist*, etc. (see, e.g., Spelke 1990); in colour perception constraints might include assumptions about the composition of common light, and so on (see, e.g., Wandell 1989). These constraints are presumed (by this author and others) to operate subpersonally in an intermediate stage of visual processing and to be at least largely impenetrable

Fitting these ideas about perceptual ambiguity into our previous discussion is straightforward and rewarding. First, we explicitly generalize the notion of a 'stimulus' to permit its application to both objective and subjective perceptual objects, that is, to intrinsic and presentational features of mind-independent objects, and to intrinsic and presentational features of sense-data. With this in hand, according to our Doctrine when I am acquainted with an ambiguous stimulus I have infallible *knowledge of* this existing thing. I do not thereby know of its essence, or know it completely or perfectly, and (at least on the nonpropositional construal) I do not know any truths about it. I simply know of this existing thing, I have a kind of epistemic anchor to it that can be exploited to formulate more or less correct descriptions of it. The reward of this application arises when we recognize that the respect in which the stimulus is ambiguous is recovered by this minimal knowledge. If S_1 is ambiguous with respect to its intrinsic shape (UR oriented cube), then in virtue of being acquainted with S_1 I do not have complete or perfect or infallible knowledge of that intrinsic shape—even if, I claim, S_1 is a sense-datum. One might argue that the knowledge I do possess of it is greater than mere knowledge of this existing thing. In virtue of being acquainted with S_1 perhaps I acquire knowledge of P_1, the "presentational shape" or "shape appearance" of S_1. While I think there is something to this (though see Footnote 24 and §3), such knowledge should probably not be deemed infallible, is difficult to make concrete, and in any case our Doctrine is sufficient to accommodate the epistemic impurities ambiguous stimuli generate. We now possess the resources to examine a challenge posed to sense-datum theory, and acquaintance knowledge generally, by perceptual constancies.

Perceptual Constancy and Perceptual Objects

The Challenge

As mentioned above one common argument for IR seeks to show that perceptual relativity, perceptual illusion, and hallucination collectively give solid evidence for the existence of sense-data as mediators of our perceptions of the mind-independent world. This has naturally led to a host of objections to such arguments,[26] one of the

(Footnote 25 continued)

by higher-level cognition (see, e.g., Raftopoulos 2009, 2010). The vision system applies them to illumination information retrieved by the retina (in early vision) and computes or "judges" which disambiguation(s) represent the most probable objects of (i.e., external objects causing) that perceptual state. What I suggest follows squarely in this framework. However, I am not committed to the variety of cognitive impenetrability that, e.g., Raftopoulos defends, but instead have sympathies with cognitive penetration (see, e.g., Macpherson 2012).

[26]I have attempted to reply to one influential objection of this sort in Brown (2012).

more recent and interesting of which can be drawn from Smith (2002). Smith argues that various perceptual cases traditionally thought to be illusions on closer inspection are not. These cases include perceptions of the famed tilted penny, Russell's perceptions of the shape of his table as he walks around it (described in *Problems of Philosophy*), among others. The ones of particular interest to Smith fail to be illusions because they are in fact instances of perceptual *constancy* (see below). The challenge is then that indirect realism cannot countenance constancy:

> The key to an answer to our Problem [of perception]…is the recognition that we are not, even in this domain [of perception], aware of perceptual sensations as objects because, *if we were, perceptual constancy would be wholly absent*: the object of awareness would appear to change whenever there was a change of sensations, because such sensations would *be* our objects. For what must a sense-datum theorist say of the typical situation in which an object is seen to approach me? He must say that the sense-datum, that which is 'given to sense,' that of which I am most fundamentally and immediately aware, *gets bigger.* But that of which I am most fundamentally and immediately aware, what is *given* to me, does not appear to change at all in such a situation. This is a plain phenomenological fact. [178]

The argument can be reconstructed as a *modus tollens*: if sensations/sense-data were perceptual objects then "constancy would be wholly absent"; constancy is present; therefore we must not have sense-data as perceptual objects. Articulating why one might accept the first premise requires some work.

Perceptual constancy requires comparisons either between perceptions or between parts of a perception. With respect to the former, when the penny rotates from being untilted to tilted relative to the perceiver and the perceiver sees it to be an object with a constant shape then constancy is present in the relevant sense. Regarding the latter, when different parts of a wall are differently illuminated and the wall is still seen to be uniformly coloured then constancy is present. More generally constancy is found in most perceptual domains, including perceived shapes, colours, sizes, and so on, and occurs when there is a perceived *constancy* in some feature of an object despite a present *variability*. What counts as a variability is domain specific: in shape perception it includes variabilities in the relative orientation of the object (e.g., the penny case), variabilities in the light transmission properties of the medium (e.g., the bent stick case), and so on; in colour perception it includes variabilities in the nature of the incident light (e.g., the wall case), of the other reflectance properties in the scene (e.g., colour constancy can occur when one looks at a scene through coloured lenses), and so on. I called these variabilities 'present' to be deliberately vague. They are generally taken to be registered by one's perceptual systems but whether or not they are consciously perceived is a subtle matter we do have not the space to debate. For example, on some views (e.g., standard interpretations of Retinex theory such as Land 1986) colour constancy is facilitated through some kind of adaptive mechanism (e.g., von Kries adaptation) whose processes may fall below the threshold of perceiver awareness. By contrast colour constancies involving partially shadowed objects can clearly contain a perceived variability (i.e., in illumination). Constancies in which what is variable is

consciously perceived—what Smith calls 'sensuous constancies'—and not merely subpersonally registered by the agent's perceptual systems are most relevant to the above argument and hence will be our focus.

Of significance is the issue of how these perceived variabilities and perceived constancies are present in, and perhaps combine in, experience. Here is where IR purportedly fails. Because of constancy the object of perception is perceived to be intrinsically unaltered; yet that object is also perceived to be altered in some more relative way.[27] This combination of factors requires that the object of perception be something we can have differing perspectives on, that for example I can perceive the shape of something in a variety of different ways or from a variety of different perspectives. This much the current criticism of IR has in common with some others, but what is done with this observation is original. Whereas others have made the weaker claim that we can understand these relativities of perception without resort to sense-data,[28] according to the current challenge sense-data cannot meet this dual demand of perceptual relativity and constancy and therefore cannot *be* the perceptual objects with which we are engaged.

The key postulate is that sense-data are not items we can have differing perspectives on. Smith believes that "there are no perspectives to be had on our sensations, and so they have no further aspects that transcend our current awareness of them. We can attend more fully to a sensation, but we cannot turn it over and contemplate its different aspects—not even in our mind's eye" (Smith 2002, 135). He goes so far as to use this to distinguish between sensation (i.e., awareness of bodily states) and perception (i.e., awareness of a world outside oneself): "Where there is the possibility of different perspectives on a single object, we have genuinely perceptual experience rather than mere sensation" (ibid).

Even if one does not wish to accept this way of distinguishing between sensation and perception, we can see why sense-data are arguably inappropriate candidates for perceptual objects. Sense-data are subjective entities, parts of oneself. You cannot get (e.g.,) several feet closer to your sense-data. So when size constancy occurs, and you see a perceptual object to be intrinsically the same in size but getting closer to you, sense-data are not the kind of thing that can meet these demands. By contrast objects in a world outside oneself straightforwardly are.

There is much that can and should be said about this argument, and many aspects of it that I have left out. Given length constraints I focus on two responses on behalf of sense-datum theory and IR, one that requires a fairly rich EF and one that does not.

---

[27]Perceptual constancies "involve a change in [e.g.,] visual experience, a change in visual sensation, despite the fact that the object of awareness does not itself appear to change at all…the changing sensations always manifest to us a changing *relation* in which an intrinsically unchanging object comes to stand to us" (Smith 2002, 172).

[28]See, e.g., Dawes-Hicks (1912, 1913/1914), Dummett (1979), Burnyeat (1979/1980), Demopoulos (2003), Schellenberg (2008), and so on.

The 'Private Space' Response

Our explanandum is the "plain phenomenological fact" that "that of which I am most fundamentally and immediately aware, what is *given* to me, does not appear to change at all in such a situation." One simple explanans posits a three-dimensional space for sense-data that is distinct from physical space.[29] Here a given sense-datum can approach an agent in her private space while maintaining a constant size within that space. In this sense a perceiver can have varying 'perspectives' on a sense-datum. This explains the "plain phenomenological fact" by reference to this object (i.e., the sense-datum) and its changing position in private space; the fact is explained in terms of the perceiver's acquaintance with this object in this space. Compelling as this explanation may seem, it also commits one to the kind of EF sense-datum advocates may wish to avoid. To see why consider by analogy the following acquaintance-based *direct* realist explanation.

A direct realist explanation in terms of acquaintance with objective things in objective space might proceed as follows: the object looks to be of constant size but approaching because the perceiver perceptually experiences (i.e., is acquainted with) precisely what is before her, an object of constant size approaching her. One might even elaborate and point out that this is so despite the fact that the projected image the object traces on the perceiver's retinas is increasing in size as the object moves toward the perceiver. The difficulty with this appeal to acquaintance is that it doesn't permit errors involving constancies, and such errors are not difficult to formulate.[30]

I suggest that shape and size constancy should generally be understood within a disambiguation framework like the one above articulated.[31] Consider the famed tilted penny as an example of shape constancy: the agent perceives the shape of the object to be constant despite the fact that we vary the orientation of the object's shape with respect to the agent (e.g., tilt the penny). When the object is tilted in this

[29]See, e.g., Russell's distinction between the "apparent" and "private" space of sense-data and the "real" and "public" space of physical things in Chap. 3 of *Problems of Philosophy*. See also Siegel (2006), who calls this general kind of response the "complex sense-data option" (391; see, esp. pp. 384–5). Note that Siegel doesn't consider the criticism of this response made in what follows.

[30]E.g., Meadows (2013) considers a different kind of error involving constancies than what is in the text. The example in the text is more directly relevant to the aims of this paper and hence what I focus on.

[31]Please note that the following analysis of shape and size constancy does not generalize to colour constancy. For a discussion of the latter see Brown (2014), where a model of colour constancy is developed that is consistent with a variety of perceptual theories, including sense-datum theory. We are here focusing on shape constancy and its peculiarities because of the use to which Smith has put it. In my view Smith's argument becomes far less plausible when applied to colour constancy phenomena (something he does not do in any detail), so avoiding discussion of the topic does not diminish from the cogency of my response.

way the set of presentational features of the stimulus is geometrically ambiguous between an elliptical object being viewed head-on (elliptical disambiguation), a round one being viewed at an angle (round disambiguation), and so on. In this case the visual system does not treat both disambiguations as equiprobable but instead favours the round one, we tend to see this stimulus as a tilted penny instead of as an untilted elliptical object. The implementation of this disambiguation favouritism is the mechanism that grounds shape constancy perception for our candidate direct realist.

There are many reasons we can offer for why this preference obtains. Perhaps most notable are evolutionary and earlier life experience, factors that have been adjusted by reference to our environment and hence have absorbed relevant contingencies. These might include the Euclidean character of local space (when considering evolutionary learning), the relative absence of elliptical objects in our environment, or the roundness and copper colour of pennies (when considering life learning), and so on. The important point for our purposes is that the preference does obtain, and that *it obtains independently of the actual objective stimulus in a given case*. If, as supposed, the object *is* a tilted penny then one's seeing it as such is accurate and hence no illusion should be ascribed. However, if in another circumstance the object is in fact elliptical and untilted—an appropriately oriented *faux penny*—then one's vision system would still prefer the round disambiguation and hence one would be prompted to incorrectly see it as a tilted penny. This misperception is arguably arising because of misleading cues, for it takes a rather special (given our environment) object being oriented in a rather specific way to prompt the misperception. One could thus argue that it is illusory.

In this respect I partly agree and partly disagree with Smith's analysis of perceptual illusion. He correctly asserts that a perception of a tilted penny does not constitute an illusory experience.[32] However, the claim is importantly limited, for perceptions of an appropriately oriented faux penny *are* illusory. Thus on a charitable reading the point of the tilted penny case has never been primarily to suggest that we typically misperceive tilted pennies, it has been to suggest that perceptual ambiguities can yield illusions.

With this in mind, what is most relevant is that consistently disambiguating stimulus ambiguities results in shape and size constancies. We see a rotating penny as having a constant, intrinsic, circular shape not because seeing it that way is *forced* on or *given* to us by presented objects in the objective world, but because that succession of ambiguous stimuli are disambiguated in accordance with the operational constraints of our vision system, constraints which prefer the "rigid rotating" disambiguation over the "non-rigid nonrotating" disambiguation. We would have the same response if we were instead seeing a nonrotating object whose shape was constantly changing in a way that mimicked the relative shape variations

[32]"[I]n no sense, not even in the extended sense given to the term in these pages, is the look of such a tilted penny an illusion" (Smith 2002, 172). On this issue Schellenberg's (2008) view is in agreement with Smith's, and thus subject to the same analysis.

of a rotating penny.[33] In short, *perceived shape and size constancy need not correspond to actual objective constancy*. The "plain phenomenological fact" that "that of which I am most fundamentally and immediately aware, what is *given* to me, does not appear to change at all in such a situation" cannot be explained simply by reference to a constant (rotating or approaching) objective thing that one is acquainted with. Instead an acquaintance-based direct realism must reference something else to explain shape and size constancy.

One may wish at this point to appeal to disjunctivism and claim that: (a) when constancy experiences are of constant objective things this is because one is acquainted with such things; and (b) when constancy "experiences" are of objective things that are not constant this is because one is in a categorically distinct non-perceptual mental state.[34] However, my aim isn't to rehearse how a direct realist who utilizes acquaintance might explain perceptual constancies, it is to consider how an indirect realist who utilizes acquaintance should. The direct realist runs into difficulties for the familiar reason that the relevant cases, shape and size constancy cases, seem subject to error, thus mandating the inadequacy of appealing *only* to one's acquaintance with objective things (and their orientations, locations, motions, etc.) to explain the cases. An analogous challenge can be posed to the indirect realist.

A purely acquaintance-based IR response requires a robust form of EF. To explain the "plain phenomenological fact" that "that of which I am most fundamentally and immediately aware, what is given to me, does not appear to change at all in such a situation," this IR asserts that the perceptual object does not change in this situation. A sense-datum perceptually appears to be round and getting closer iff it is round and getting closer. This is how appeal to a three-dimensional private space generates an explanation of shape and size constancy experiences. The cost, however, is that a rich EF holds: there can only be a kind of perfect perceptual awareness of the relevant shapes, sizes and motions of sense-data (in these cases).

Some indirect realists may feel perfectly comfortable with this position. However, as mentioned above, there are important reasons to not rely on this kind of EF. Suppose I am acquainted with an elliptical, upright sense-datum. Why is it, not merely improbable, but *impossible* for this object to seem like a tilted round thing? One can operationally define acquaintance with sense-data in such a way as to make this impossible (which is equivalent to asserting EF about these features), but a critic will argue that there isn't a compelling, non-question-begging reason to do so. By contrast, if we allow for the possibility of an elliptical, upright sense-datum to perceptually seem round and tilted, much as we allow for an elliptical, upright objective *faux penny* to perceptually seem round and tilted, then merely postulating

[33]Consider the fact that the shapes on a television screen that represent a rotating penny are not themselves constant but instead constantly changing. We do not, however, see them as representing a nonrigid nonrotating object but instead as representing a rigid rotating object. This achievement is due to a disambiguation bias in our vision systems, not due to an intrinsic roundness being forced upon us.

[34]E.g., Byrne and Logue (2008) is a recent collection on disjunctivism.

that sense-data can be variously positioned in a three-dimensional private space is inadequate to explain the full gamut of shape and size constancy experiences within IR. Consider, therefore, a different explanation.

The 'Two-Factor' Response[35]

Holding EF for one's experience of sense-data typically involves asserting that something perceptually appears *F* to perceiver *A* iff that thing—which is a sense-datum—*is F* (see §1, esp. Footnote 2). To the typical sense-datum theorist the generality of this inference as stated renders it implausible. For example if something perceptually appears to be a tree to perceiver *A* then it would follow that the sense-datum being experienced is a tree, a problematic conclusion. One familiar means of avoiding such conclusions places a limit on when this inference is applicable. The rough idea is that the inference is applicable for members of a set of features which collectively define the *sensory core* of a perceptual experience, and the inference is potentially not applicable for all other features, these latter being in the *sensory periphery* (e.g., Price 1932). The core might for example involve perceptually basic or low level properties like shapes and colours, while the periphery might contain less basic properties like being a tree or a cup. Since the inference from '*x* appears *F*' to '*x* is *F*' holds of the core features, one's perceptual experiences of these features of sense-data can be explained via acquaintance and this can define a limited kind of EF. By contrast since the inference from '*x* appears *F*' to '*x* is *F*' needn't hold of the peripheral features, perceptual experiences of these features *cannot* be explained via only acquaintance with sense-data and their features. Instead, appeal to a fallible mechanism like the description/representation of sense-data is required. Because sense-data can be represented to generate experiences of peripheral features we can in this sense have differing 'perspectives' on them.[36] The result is a *two-factor* acquaintance and representation approach perceptual experience.

To apply this framework to our case study of shape and size constancy we put '*x* appears to be of a constant size and approaching' in the sensory periphery instead of the core. By virtue of being in the periphery these constancy experiences would involve representations and hence allow for at least some of them to be erroneous. This overcomes the weakness of the private space response. However, it is worth developing this two-factor response more fully.

[35]Siegel (2006) considers a two-factor response to Smith's challenge. There are a few differences in our presentation that I will leave to the reader, and an important difference in analysis: she does not consider what's left in the sensory core once the two-factor response has been employed and what consequences may follow from that consideration; by contrast in what follows I focus precisely on these issues.

[36]I presume that we can remain mute as to whether or not these representations involve concepts.

Since on this analysis constancy aspects of shape and size experiences are in the sensory periphery, it is natural to wonder what is left in the sensory core, particularly with regard to geometric features. The shapes and sizes of sense-data were originally supposed to be in the core, not the periphery, permitting perceivers to have some fairly concrete foundational knowledge about what is perceptually experienced. Without the shapes and sizes of sense-data in the core, that foundation dissipates. Can any kind of foundational geometric knowledge of what is perceived be recovered?

One suggestion is to leave *intrinsic* geometric features like shape and size out of the core and instead identify more relative geometric features like shape and size *appearances* for the core. While this option is worth considering in detail, I wish to identify three immediate concerns and then move to my preferred option. The first is a concern over what these geometric appearance properties of sense-data actually are. One possible answer is that they are analogous to the projected retinal images (see again Hatfield and Epstein 1979), introducing the somewhat unattractive need for perceptual awareness of projective features of sense-data and (perhaps) some means of such projection occurring. Second, as Smith points out, we are very poor at identifying anything like geometric appearance properties in perception (2002, 181–2; 2006, 416), making such properties odd candidates for being in the sensory core. Finally, and perhaps most importantly, if constant sizes and shapes of sense-data are in the sensory periphery, playing second fiddle to the shape appearance core, the resulting account serves as a very poor explanans for our explanandum, that "that of which I am most fundamentally and immediately aware, what is given to me, does not appear to change at all in such a situation."

The alternative I propose is that, at least for now, we leave in the sensory core only the minimal acquaintance knowledge outlined in our Doctrine: knowledge of that existing thing. This has the effect of massively shrinking, indeed almost obliterating, any simple EF for sense-datum theory and IR more generally. However, it avoids the above worries and permits apparent shape and size constancies to be "plain phenomenal facts" arising not merely through being acquainted with objects of constant shape and size, but through "plainly" representing these objects as maintaining constant shape and size. By contrast actual shape and size constancy will be determined by what the relevant sense-data are doing (for direct perception) and what any objective things represented by these sense-data are doing (for indirect perception). With regard to sense-data, this proposal is compatible with sense-data being in a three-dimensional private space, a projected two-dimensional one, or whatever.

In slightly more detail, and accordance with §1, on this proposal the indirect realist has to accept that the intrinsic shape and size of a sense-datum is not forced on a perceiver by being aware of or acquainted with it. Instead, what is via acquaintance epistemically given or available to a perceiver is ambiguous, and when that object "appears to", in Smith's sense, have some intrinsic shape and size, this "appearing to" is accomplished by disambiguating this given via some capacity for description or representation. Thus, in Smith's scenario, the sense-datum may indeed get larger, and hence the agent is acquainted with a sense-datum of

increasing size. However, what this *alone* makes epistemically available or gives to the agent is ambiguous between this sense-datum increasing in size and retaining a constant distance, or approaching the agent and retaining a constant size. For the sense-datum to appear one of these ways to the exclusion of the other requires disambiguating this given. Acquaintance alone cannot achieve this. Something in addition, namely description or representation, is needed.

There are many questions that two-factor theorists of all stripes must answer. Do these perceptual representations utilize concepts and if so what kinds of concepts? What connection do these representations have to thoughts and cognition? What is the contribution of each of acquaintance and representation to perceptual experience? How are these two-factors "united" in perceptual experience? And so on. Smith for example considers two-factor theories in which the perceptual representation involves concepts (what he calls 'dual-component theories') at some length in his wonderful book and launches several criticisms at them (2002, 67–93). The criticisms are valuable but tangential to understanding perceptual constancies and thus demand separate treatment. Consider finally three queries about the present proposal and a brief remark about each.

1. Is the two-factor response really preferable to the three-dimensional private space response? The answer largely depends on how strongly one feels that the sense-datum theorist should permit the possibility of experiential error for the shapes, sizes, locations and orientations of sense-data. As stated above, I know of no compelling reason to exclude these possibilities—Why can't a stationary but growing sense-datum appear to be of constant size and moving toward me?—and doing so by stipulation will surely strike some as unsatisfactory. Stated more generally, the answer depends on how strongly one is tied to EF. As I have stated throughout, strong ties to EF have created unnecessary problems for IR, and it is worth having an IR available that avoids them.
2. Why choose sense-datum theory and IR generally instead of a direct realism of either a purely acquaintance or two-factor sort? This is a broad, difficult question. Let me address a more focused one with a similar sentiment. If we introduce the possibility of illusion for basic sensory features of sense-data like shapes and sizes, then we lose a central motive for IR, namely EF: that there are basic aspects of what is experienced in perception (e.g., shapes, sizes, colours, etc.) that *we cannot be wrong about*. But without EF sense-datum theory loses its main advantage over direct realism, and what's left is an epistemology ($\sim$EF), account of illusion (representationalist), and so on that arguably look much like they do in direct realism. So why add sense-data between the perceiver and the objective world?

   This is a challenging problem for IR. Here is my short answer. (A) I suspect that there are fundamentally different kinds of illusions (see, e.g., Brown 2012): in some cases the erroneous/illusory aspects are best understood by reference to misrepresenting the perceptual object (in which case the erroneously ascribed

features are not instantiated by the perceptual object); in some cases the erroneous aspects are best understood by reference to a perceptual object that instantiates those features (in which case the object or at least the feature is not objective). Something like the Argument from Illusion is compelling for illusions of the second sort, but not the first. We only need the Argument from Illusion to work for a decent number of illusions to ground IR and I suspect that there are a decent number of illusions of the second sort. (B) When we move beyond illusions and consider hallucinations, other perceptual relativities (e.g., normal variations in colour perception across individuals and species), the lack of objective correlates for central features of sensory qualities like the hues, saturations and lightnesses of colour (Hardin 1988), and so on, we realize that there remain plenty of reasons to endorse IR. Given the present proposal, arguments for IR cannot appeal to a rich EF. However, the many perceptual phenomena that have been used to ground IR remain extremely vexing, even without a rich EF. Thus, one might find for example that the best explanation of these phenomena is IR, regardless of whether or not one endorses a robust EF.

3. Without EF, isn't the indirect realist epistemically lost? Suppose there remains adequate reason to be an indirect realist. On the current view we have lost all EF that stretches past knowledge *of that existing thing*. So now we're stuck with the perceptual veil (IR) and no robust EF with which to build a solid epistemology.

I appreciate that this looks glum. I am more optimistic, seeing this as a reasonable location from which to build an informative perceptual epistemology. Perhaps through such efforts some form of EF will re-emerge, in which case the current proposal is merely a basecamp from which a stronger foundation of basic propositions for perceptual epistemology can be uncovered. Alternatively, perhaps the whole notion of basic perceptual knowledge that goes beyond knowledge *of that existing thing* is unrecoverable. In this case we need to either build our perceptual epistemology without such basic foundations, or give into skepticism. This strikes me as a healthy position from which to work.

Conclusion

According to the deflated epistemic foundationalism about perception under consideration (§1), if acquaintance with a thing, be that an objective or subjective thing, is taken to constitute or immediately yield basic propositional knowledge of the thing, that knowledge is extremely limited. It should not be taken to consist of knowledge of the thing's shape or size or (I would argue) colour, et cetera. These features, insofar as they are intrinsic to a perceived thing, need not make their natures "manifest" by virtue of being presented to one in perception. They instead may often become epistemically salient to one through one's perceptual descriptions or representations of what is presented, which are invariably fallible. This yields a two-factor (acquaintance-representation) approach to perception.

I tried to illustrate how we should broadly conceive of sense-datum theory when it is divorced from epistemic foundationalism (§1), and applied this conception to shape and size constancy perceptions in an effort to dissolve challenges these constancies might be thought to pose for sense-datum theory. That application proceeded in two steps. The first was a re-examination of the roots of ambiguous objective stimuli (e.g., the wire cube) and perceptual ambiguity (e.g. seeing the cube at one time as a UR cube and the next as a DL one). The second used this as a framework for understanding why we could accept the possibility of error in shape and size constancy experiences, even when focused on one's acquaintance with one's own sense-data. It is antecedently reasonable to suppose that we can experience an elliptical upright sense-datum as a round tilted object, thus laying the groundwork for an account of constancy that relies on (but does not require) a two-dimensional array of sense-data. Such an account necessitates a conception of sense-datum theory with a deflated epistemic foundationalism of the above sort, but in any case is no inherent threat to the theory. Two alternatives to this picture were considered, both of which demand a richer foundationalism.

One alternative seeks to explain shape and size constancy experiences purely by appeal to acquaintance (i.e., perceptual contact). This requires a kind of epistemic foundationalism pertaining to the shapes and sizes of perceptual objects (of whatever kind) that I find difficult to justify. Perhaps that justification has simply eluded me, in which case a purely acquaintance-based view whose perceptual objects are in some kind of three-dimensional space is adequate to explain shape and size constancies. Either way, we should admit that we can have differing "perspectives" on all candidate perceptual objects, sense-data or objective things, although care may be needed in the term's explication in a given application.

The acquaintance theorist who concedes that acquaintance with an object need not yield infallible knowledge of its intrinsic shape and size may still demand more than our deflated Doctrine. She may for example believe that in addition to knowledge of an existing thing, acquaintance yields infallible knowledge of a perceived thing's *presented* shape and size, that is, of its shape and size *appearances*. For example, although in virtue of being acquainted with some objective thing or sense-datum I may not thereby know that it is intrinsically elliptical, I nonetheless do thereby know that its presented shape is elliptical, I know its "shape appearance". This final alternative is a tempting line of thought, but one that I argued is the least attractive. We don't seem to have very firm propositional knowledge of shape and size appearances, and even if we did they are a poor pillar from which to explain the prominence of intrinsic shapes and sizes in constancy experiences (§3). In addition, by virtue of endorsing a rich epistemic foundationalism the "appearance view", like the "three-dimensional view" conflicts with a natural reading of the role of acquaintance in perceptual knowledge (§1).

On this natural reading the purpose of acquaintance knowledge is not itself to "give" the agent various facts about what she is perceiving, it is to give her the access to what she is perceiving that is needed to formulate, through representation/description, propositions about such facts. Acquaintance provides for a perceiver an epistemic anchor to the existing intentional objects of one's

perceptual state, one that she can exploit to acquire knowledge of various truths (i.e., knowledge by description) about those things. Acquaintance theorists who insist that by being acquainted with a perceptual object we thereby have infallible knowledge of its intrinsic or presented shape and size are missing this point. From this perspective they are therefore not only unnecessarily extending their epistemic presuppositions, they are misconceiving what I believe to be the fundamental structure of perceptual knowledge the acquaintance/description division seeks to provide.

The point is not so much to mandate the deflated epistemic foundationalism for sense-datum theory as it is to isolate a conceptual location from which sense-datum theorists and adherents of indirect realism more generally can examine distinct perceptual phenomena to assess whether or not they *should* endorse foundationalism for such phenomena. That is, an indirect realist with a deflated foundationalism has the space to explore the extent to which she should adhere to foundationalism, instead of the requirement to adhere to foundationalism as much as possible. In my view the same is additionally true of all presentational or relationalist approaches to perception, whether the presented entities are sense-data or objective things.

Two matters remain particularly pressing. The sense-datum theorist must ensure that arguments for her view do not require epistemic foundationalism where it is not warranted. In addition a workable two-factor theory of experience must be developed, specifically one that contains the flexibility to endorse and reject foundationalism as needed. One key aspect of this involves examining how acquaintance knowledge and perceptual presentation are exploited to acquire descriptive knowledge, for the former pair likely contain constraints governing the latter. In this regard our conception of the *amount* of knowledge acquaintance yields may grow. But care is necessary, for the *form* of that knowledge may well be logically functional or behaviourally dispositional. It need not be propositional.

References

Armstrong, D. (1961). *Perception and the Physical World*. New York, USA: Routledge.

Austin, J. (1962). *Sense and sensibilia.* Reconstructed by G. J. Warnock. Oxford, UK: Oxford University Press.

Ayer, A. (1940). *The foundations of empirical knowledge*. London: Macmillan.

Block, N. (2003). Mental paint. In M. Hahn (Ed.), *Reflections and replies: Essays on the philosophy of Tyler Burge* (pp. 165–200). Cambridge, MA: Bradford Book.

Brewer, B. (2011). *Perception and its objects*. New York, USA: Oxford University Press.

Brown, D. (2010). Locating projectivism in intentionalism debates. *Philosophical Studies, 148*, 69–78.

Brown, D. (2012). Losing grip on the world: From illusion to sense-data. In A. Raftopoulos & P. Machamer (Eds.), *Perception, realism and the problem of reference* (pp. 68–95). Cambridge, UK: Cambridge University Press.

Brown, D. (2014). Colour constancy and colour layering. *Philosophers' Imprint, 14*(16), 1–31.

Brown, D. (Forthcoming). Projectivism and phenomenal presence. To appear in F. Macpherson & F. Dorsch (Eds.), *Phenomenal presence*. Oxford, UK: Oxford University Press.
Burge, T. (2010). *Origins of objectivity*. Oxford, UK: Oxford University Press.
Burnyeat, M. (1979/1980). Conflicting appearances. *Proceedings of the British Academy, 65*, 69–111.
Byrne, A. (2001). Intentionalism defended. *Philosophical Review, 110*, 49–90.
Byrne, A., & Logue, H. (2008). *Disjunctivism: Contemporary readings*. Cambridge, MA: MIT Press.
Campbell, J. (2002). *Reference and consciousness*. Oxford, UK: Oxford University Press.
Chalmers, D. (2006). Perception and the fall from Eden. In T. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 49–125). Oxford, UK: Oxford University Press.
Chirimuuta, M. (2008). Reflectance realism and colour constancy: What would count as scientific evidence for Hilbert's ontology of colour? *Australasian Journal of Philosophy, 86*(4), 563–582.
Chisholm, R. (1942). The problem of the speckled hen. *Mind, 51*, 368–373.
Chisholm, R. (1957). *Perceiving: A philosophical study*. Ithaca, NY: Cornell University Press.
Cohen, J. (2009). *The red and the real: An essay on colour ontology*. Oxford, UK: Oxford University Press.
Crane, T. (2006). Is there a perceptual relation? In T. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (pp. 126–146). Oxford, UK: Oxford University Press.
Dawes-Hicks, G. (1912). The nature of sense-data. *Mind, 21*, 399–409.
Dawes-Hicks, G. (1913/1914). Appearances and real existence. *Proceedings of the Aristotelian Society*, 1–48.
Demopoulos, W. (2003). Russell's structuralism and the absolute description of the world. In N. Griffin (Ed.), *Cambridge companion to Russell* (pp. 392–419). Cambridge, UK: Cambridge University Press.
Dretske, F. (1993). Conscious experience. *Mind, 102*, 263–283.
Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
Dummett, M. (1979). Common sense and physics. Reprinted in (1993) *Seas of language* (pp. 376–410). Oxford, UK: Oxford University Press.
Fumerton, R. (2005). Speckled hens and objects of acquaintance. *Philosophical Perspectives, 19*, 121–139.
Fumerton, R. (2009). Markie, speckles, and classical foundationalism. *Philosophy and Phenomenological Research, 79*, 207–212.
Gert, J. (2010). Color constancy, complexity, and counterfactual. *Noûs, 44*(4), 669–690.
Gupta, A. (2006). *Empiricism and EXPERIENCE*. New York, USA: Oxford University Press.
Hardin, C. L. (1988). *Color for philosophers: Unweaving the rainbow*. Cambridge, MA: Hackett Publishing.
Harman, G. (1990). The intrinsic quality of experience. In J. Tomberlin (Ed.), *Philosophical perspectives* (Vol. 4, pp. 31–52). Atascadero: Ridgeview Publishing Co.
Hatfield, G., & W. Epstein (1979). The sensory core and the Medieval foundations of early modern perceptual theory. *Isis 70*, 363–48. Reprinted in G. Hatfield (2009) *Perception and cognition: Essays in the philosophy of psychology* (pp. 358–385). New York, USA: Oxford University Press.
Hilbert, D. (2004). Hallucination, sense-data and direct realism. *Philosophical Studies, 120*, 185–191.
Hilbert, D. (2005). Color constancy and the complexity of color. *Philosophical Topics, 33*, 141–158.
Hopkins, R. (1998). *Picture, image and experience: A philosophical inquiry*. Cambridge, UK: Cambridge University Press.
Huemer, M. (2001). *Skepticism and the Veil of perception*. Lanham, Md: Rowman & Littlefield.

Jackson, F. (1977). *Perception: A representative theory*. Cambridge, UK: Cambridge University Press.
Jagnow, R. (2010). Shadow-experiences and the phenomenal structure of colors. *Dialectica, 64*, 187–212.
Kalderon, M. E. (2008). Metamerism, constancy, and knowing which. *Mind, 117*, 935–971.
Land, E. (1986). Recent advances in Retinex theory. *Vision Research, 26*, 7–21. Reprinted in A. Byrne & D. Hilbert (Ed.), *Readings in color, vol. 2: The science of color* (pp. 143–160). Cambridge, USA: MIT Press.
Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of and indirect mechanism. *Philosophy and Phenomenological Research, 84*(1), 24–62.
Macpherson, F., & Platchias, D. (2013). *Hallucination*. Cambridge, USA: MIT Press.
Matthen, M. (2010). How things look (And what things look that way). In B. Nanay (Ed.), *Perceiving the world* (pp. 226–253). New York, USA: Oxford University Press.
Maund, B. (2012). Perceptual constancies: Illusion and veridicality. In C. Calabi (Ed.), *Perceptual illusions: Philosophical and psychological essays* (pp. 87–106). New York, USA: Palgrave Macmillan.
McDowell, J. (1994). *Mind and world*. Cambridge, USA: Harvard University Press.
Meadows, P. (2013). On A. D. Smith's constancy based defence of direct realism. *Philosophical Studies, 163*, 513–525.
Nanay, B. (2009). How speckled is the hen. *Analysis, 69*, 499–502.
Peacocke, C. (1983). *Sense and content: Experience, thought, and their relations*. Oxford, UK: Clarendon Press.
Poston, T. (2007). Acquaintance and the problem of the speckled hen. *Philosophical Studies, 132*, 331–346.
Price, H. H. (1932). *Perception*. London, UK: Methuen.
Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neuroscience inform philosophy?*. London, UK: MIT Press.
Raftopoulos, A. (2010). Ambiguous figures and representationalism. *Synthese,*. doi:10.1007/s11229-010-9743-1.
Robinson, H. (1994). *Perception*. New York, USA: Routledge Press.
Russell, B. (1913). The nature of sense-data—A reply to Dr. Dawes Hicks. *Mind, 22*, 76–81.
Russell, B. (2007). *Problems of philosophy*. New York, USA: Cosimo.
Schellenberg, S. (2008). The situation-dependency of perception. *Journal of Philosophy, 105*, 55–84.
Siegel, S. (2006). Direct realism and perceptual consciousness. *Philosophy and Phenomenological Research, 73*(2), 378–410.
Smith, A. D. (2002). *The problem of perception*. Cambridge, USA: Harvard University Press.
Smith, A. D. (2006). In defence of direct realism. *Philosophy and Phenomenological Research, 73*(2), 411–424.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 29–56.
Stoljar, D. (2004). The argument from diaphanousness. In M. Ezcurdia, R. Stainton, & C. Viger (Eds.), *New essays in the philosophy of language and mind: Special issue of the Canadian Journal of Philosophy* (pp. 331–391). Calgary, Canada: University of Calgary Press.
Tye, M. (2000). *Consciousness, color, and content*. Cambridge, USA: MIT Press.
Tye, M. (2009). A new look at the speckled hen. *Analysis, 69*, 258–263.
Tye, M. (2010). Up close with the speckled hen. *Analysis, 70*, 283–286.
Wandell, B. A. (1989). Color constancy and the natural image. *Physica Scripta*, *39*, 187–92. Collected in A. Byrne & D. Hilbert (Eds.) (1997), *Readings on colour, vol. 2: The Science of Colour* (pp. 161–176). Cambridge, USA: MIT Press.

Author Biography

Derek H. Brown is an Associate Professor with a Research Appointment in the Department of Philosophy at Brandon University in Canada. He works in philosophy of mind and epistemology with particular interest in philosophy of colour and indirect realist approaches to perception. His work has appeared in journals such as *Philosophers' Imprint, Philosophical Quarterly,* and *Philosophical Studies*, and has appeared or is forthcoming in collections published by Cambridge University Press and Oxford University Press among others. He co-edited (with M. Frappier and R. DiSalle, 2012, Springer) *Analysis and Interpretation in the Exact Sciences: Essays in Honour of William Demopoulos*, and is currently co-editing (with F. Macpherson) *The Routledge Handbook on Philosophy of Colour*. He has held visiting appointments in philosophy at the University of Pittsburgh and the University of Glasgow.

Whitehead *Versus* Russell

Gregory Landini

Introduction

In his autobiography Russell speaks of geometry as his first love. It brought him a scholarship in 1890 to study mathematics at Trinity College, Cambridge. Alfred North Whitehead, a mathematician there, had been impressed by his scholarship examination, and took an interest in him. Their relationship would mature into a life-long friendship and collaboration on many projects including *Principia Mathematica*. It is worth quoting Russell on the subject of their friendship:

> In England, Whitehead was regarded only as a mathematician, and it was left to America to discover him as a philosopher. He and I disagreed in philosophy, so that the collaboration was no longer possible, and after he went to America I naturally saw much less of him.[1]

The word "philosophy" here is used oddly. Mathematical logic is surely part of philosophy especially for Russell's logical atomism which marks logic as the essence of philosophy. What were the philosophical differences between Whitehead and Russell and what antecedents (if any) did they have in *Principia*?

Russell goes on in his comments on Whitehead to offer a diagnosis of the source of their lack of collaboration in later years. He wrote:

> We began to drift apart during the First World War when he completely disagreed with my pacifist position. In our differences on this subject he was more tolerant than I was, and it was much more my fault than his that these differences caused a diminution in the closeness of our friendship.[2]

[1] Russell (1952).
[2] Ibid.

G. Landini (✉)
University of Iowa, Iowa City, USA
e-mail: gregory-landini@uiowa.edu

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_6

The causes of the "drifting apart" are complicated. Naturally, there were human factors. Russell explained (*op cit*):

> In the last months of the war his younger son, who was only just eighteen, was killed. This was an appalling grief to him, and it was only by an immense effort of moral discipline that he was able to go on with his work. The pain of this loss had a great deal to do with turning his thoughts to philosophy and causing him to see ways of escaping from belief in a merely mechanistic universe. His philosophy was very obscure, and there was much in it that I never succeeded in understanding. He had always had a leaning towards Kant, of whom I though ill, and when he began to develop his own philosophy he was considerably influenced by Bergson. He was impressed by the aspect of unity in the universe, and considered that it is only through this aspect that scientific theories can be justified. My temperament led me in the opposite direction, but I doubt whether pure reason could have decided which of us was more nearly in the right. Those who prefer his outlook might say that while he aimed at bringing comfort to plain people, I aimed at bringing discomfort to philosophers; one who favored my outlook might retort that while he pleased the philosophers, I amused plain people. However that may be, we went our separate ways, thought affection survived to the last.[3]

These comments, it seems to me, are unfair to Whitehead. Russell was all too eager to engage in a polemic about the importance of freeing the mind from falling into the trap of fashioning a metaphysic to suit personal hopes and fears. The "essence of religion," in Russell's view, is the objective rational contemplation and reverence for the universe *sub specie aeternatitis*. Russell is justified. But we shall see that Whitehead's view about nature being unified was no different from that invoked by Einstein in his theory of general relativity.

Russell was undoubtedly chagrined that *Principia*'*s* volume IV on geometry never appeared. The root causes of the volume IV of *Principia* never appearing, and the exact time of its abandonment by Whitehead (if indeed there ever was an official abandonment), remains unclear. All the work notes were destroyed after his death in 1947. The collapse of *Principia's* volume IV on geometry was surely a disappointment to Russell. He had the impression that by 1914 Whitehead had a bulk of the work finished, at least in outline. Indeed, the 11th edition for 1910–11 of the *Encyclopedia Britannica* hosted Whitehead's papers "The Axioms of Geometry," and "Non-Euclidean Geometry." But we shall find that the challenge of Einstein's work on general relativity altered many things. Russell seems not to have noticed a connection. He wrote:

> Whitehead was to have written a fourth volume, on geometry, which would have been entirely his work. A good deal of this was done, and I hope it still exists. But his increasing interest in philosophy led him to think other work more important. He proposed to treat a space as the field of a single triadic, tetradic, or pentadic relation*, a treatment to which, he said, he had been led by reading Veblen. *And generally a space of n dimensions as the field of an $(n + 1)$ relation.[4]

Whitehead's investigations on eliminating *points* in geometry, which Russell so praised in his 1914 *Our Knowledge of the External World*, seem to have led him to

[3]Ibid.

[4]Russell (1948).

engage Einstein's views by developing a *method of extensive abstraction*. It appears prominently in his 1919 book *The Concept of Nature*. Whitehead's work on extensive abstraction became part of his conception of *pure* geometry.

The method of extensive abstraction also appear in Whitehead's 1925 book *The Principles of Natural Knowledge*, dedicated to his son Eric who had died in 1918 in the first world war. Whitehead became captivated by the idea of finding principle of *general relatedness* that would rival that of Einstein's principle of general relativity. Influenced by Eddington and Mach, we find that Russell accepted Einstein's positions *tout court*. Russell's (1927a) *The Analysis of Matter* shows little sympathy for Whitehead's work. Russell had become engaged head long in his physicalism—a version of neutral monism according to which matter and mind do not exist. Matter is a fiction that lives in series of events (transient physical particulars) obeying the new relativistic laws. Mind is also a fiction, a series of event (transient physical particulars) obeying largely behavioristic laws. In contrast, Whitehead eventually worked out a conception of events as *processes* that regards nature and mind as organically unified in way that makes Russell's neutral monism impossible.

By 1929 and the appearance of Whitehead's *Process and Reality*, the two had come far apart. But there are many more differences that show up earlier than might have at first been imagined. *Principia* itself displays some of the clashes. Volume II, which was originally to appear in 1911 was delayed until 1912 by Whitehead who was *solely* responsible for an extensive emendation of the volume. The problem was that unruly ambiguous expressions were allowed that could change their meaning in repeated occurrence in a given proposition. For example, recall that a class σ is *similar* to α (written as $\sigma\ sm\ \alpha$) if and only if all the members of σ can be put into one-to-one correlation with all the members of α. But nothing in the notation "$\sigma\ sm\ \alpha$" notation tell us whether σ is the same relative type of class as α, or whether it is of higher relative type or lower. Whitehead explained the unruly ambiguity as follows (Whitehead and Russell 1957, vol. II, p. 11):

> When a typically ambiguous symbol such as "*sm*" or "Nc" occurs more than once in a given context, it must not be assumed, unless required by the conditions of significance, that it is to receive the same typical determination in each case. Thus, e.g., we write "α *sm* β.⊃.β *sm* α," although if α and β are of different types the two symbols "*sm*" must receive different typical determinations.

This matter does not concern types of *individuals*. It concerns *relative types* of classes couched within *Principia's* no-classes theory. We shall find that some of Whitehead's amendments are very bad. And judging from a few letters remaining of their 1911 correspondence that Russell kept, some of Whitehead's amendments seemed unsatisfactory to Russell, especially Whitehead's new convention IIT of *126 which changes the way typical ambiguities had been understood in the entire previous part of the book.

The problems are entirely fixable.[5] But neither Whitehead nor Russell took the time to fix them, not even in the 1925 second edition. One wonders why Whitehead's bad amendments were not addressed there. Victor Lowe reports that a letter of 24 May 1923 was found in in the papers of Dora Russell that suggests that there was at least some collaboration on the second edition. Part of the letter reads as follows:

> I don't think that "Types" are quite right. They are "tending towards the truth," as the Hindoo said of his fifth lie on the same subject. But for heaven's sake, don't alter them in the text.[6]

Lowe thinks that Whitehead is anticipating what would become by the 1940s a widely touted criticism the *Principia's* theory of types of *individuals*. But more likely, Whitehead and come to realize that some of his 1911 emendations for volume II on relative types of classes were not quite right.

The second edition Heralds Russell's new section *8 which sets out a system of deduction for quantification theory without free variables. It is the first of its kind, anticipating Quine's work by some fifteen years. But apart from *8, nothing Russell put into the second edition was intended to be endorsed. Sadly, this has misled readers for decades. Russell had put into the second edition an evaluation of a formalization of some experimental ideas from Wittgenstein's *Tractatus*. Whitehead may well have considered such an evaluation entirely out of place in the work.

In this paper we hope to sort out some of the intellectual differences between Whitehead and Russell that are relevant to *Principia*, including is proposed volume IV on geometry. We shall take up the issues of typical ambiguity, the nature of classes, geometry, and the existence of mind and matter.

The Logic of Relations and the Question of Order

Russell and Whitehead held that mathematics is a science of order. It studies all the kinds of structures that there are by studying the way relations order their fields. Mathematics, including, the non-Euclidean geometries, consists of the logic of relations. In 1900, shortly after they attended a congress in Paris devoted to mathematics, Russell wrote "On the Logic of Relations." It was translated into French and published in Peano's journal *Revista di Mathematica* in 1901. It is important to understand that Russell's logic of relations is not to be identified with polyadic (as opposed to monadic) quantification theory. When Russell speaks of the "logic of relations," he means to imply that the *impredicative comprehension* of relations (and properties) in intension is part of (or emulated by) logic. Let me call this ^{cp}Logic. This is a radical departure from polyadic quantification theory because it allows existence theorems in pure logic—theorems governing the existence of

[5] See Landini (2015).

[6] Quoted from Lowe (1990, p. 276).

properties and relations. It is impredicative comprehension that *alone* makes logic an informative (synthetic) yet *a priori* science. Quantification theory, whether first or higher order, is not an informative science—though there is no decision procedure for its logical truths.

Naïve comprehension yields paradoxes (of attributes and of classes/relation-in-extension) which Russell discovered in 1901. After a long struggle, Russell felt that although we are acquainted with many attributes, it is wholly impossible to find purely logical principles to discern which *wffs* comprehending attributes are safe and which yield contradictions. A simple type regimentation blocks the paradoxes, but Russell felt that types of entities to be out of sorts with the nature logic. By 1905, Russell came to believe that logic must *emulate* a formal system for the simple-type regimented comprehension of attributes in intension. After a diligent effort to emulate[7] the formal grammar of simple type regimented predicate variables, Russell and Whitehead finally settled on the formal language of *Principia Mathematica*.

The formal language of *Principia* adopts schematic letters φ, ψ, f and g for *wffs* of its object language and under conventions of typically ambiguity it adopts $\varphi!$, $\psi!$, $f!$ and $g!$, etc., as bindable predicate variables. It was the first system ever to explicitly adopt axiom schemata of *impredicative*[8] comprehension. For example, we find:

$$^{*}12.1\ (\exists f)(\varphi x \equiv_{x} f!x)$$
$$^{*}12.11\ (\exists f)\big(\varphi xy \equiv_{x,y} f!x, y\big).$$

If we were to restore modern simple type indices to the predicate variables, the shriek would no longer be needed. We would have (respectively) the following:

$$\left(\exists f^{(t)}\right)\left(\varphi(x^{t}) \equiv_{x^{t}} f^{(t)}(x^{t})\right)$$

$$\left(\exists f^{(t_1,t_2)}\right)\left(\varphi(x^{t_1}, y^{t_2}) \equiv_{x^{t_1},y^{t_2}} f^{(t_1,t_2)}(x^{t_1}, y^{t_2})\right).$$

To this day, readers of *Principia* are astonished to find Whitehead and Russell so well ahead of their time. Some still reject the interpretation that there are schematic letters in *Principia.* But without them the deductive theory becomes unintelligible.

Emulating (or embracing) the impredicative comprehension of attributes in intension does not tell us how to form a theory of classes or relations-in-extension. But with the scope distinctions afforded by Russell's 1905 theory of descriptions and his new ^{cp}Logic, Russell was able to introduce extensional contexts and thus

[7]This was Russell's substitutional theory of propositional structure. It was based on Russell's 1905 theory of definite descriptions and required an ontology of propositions as objective truths and falsehoods. It lasted roughly from 1905 through 1907 and perhaps even into 1908 and appears in Russell's paper "Mathematical Logic as Based on the Theory of Types."

[8]What makes this axiom schema *impredicative* is that the schematic letter φ stands for any *wff* no matter whether it contains bound predicate variables.

develop an adequate no-classes and no-relations in extension theory. For classes of individuals, *Principia* has the following:

$$[\hat{z}(\psi z)][f\{\hat{z}(\psi z)\}] = df_{*20.01}(\exists\varphi)(\varphi!x \equiv_x \psi x .\&. f\{\varphi!\hat{z}\}).$$
$$x \in \varphi!\hat{z} = df_{*20.02}\varphi!x.$$

This does not emulate classes of classes of individuals. For that *Principia* has further definitions:

$$(\alpha)f\alpha = df_{*20.07}(\varphi)(f\{\hat{z}(\varphi!z)\})$$
$$(\exists\alpha)f\alpha = df_{*20.071}(\exists\varphi)(f\{\hat{z}(\varphi!z)\})$$
$$[\hat{\alpha}(\psi\alpha)][f\{\hat{\alpha}(\psi\alpha)\}] = df_{*20.08}(\exists\varphi)(\varphi!\alpha \equiv_\alpha \psi\alpha .\&. f\{\varphi!\hat{\alpha}\}).$$
$$\beta \in \varphi!\hat{\alpha} = df_{*20.081}\varphi!\beta.$$

These do not emulate classes of classes of classes of individuals. But we can readily see the pattern. Indeed, we can see that there are two sorts of class expressions in *Principia,* namely, $\hat{z}(\psi z)$ and $\hat{\alpha}(\psi\alpha)$. Free lower-case Greek α, β, γ, δ, η, σ, ρ etc., stand in for class expressions of one or another of these forms.

For relations in extension of individuals, *Principia* offers analogous definitions. For instance, we find the following:

$$[\hat{x}\hat{y}(\psi xy)][f\{\hat{x}\hat{y}(\psi xy)\}] = df_{*21.01}(\exists\varphi)\big(\varphi!xy \equiv_{x,y} \psi xy .\&. f\{\varphi!\hat{x}\hat{y}\}\big).$$
$$z\{\hat{x}\hat{y}(\varphi!xy)\}w = df_{*21.02}\ \varphi!zw$$

The second is a stipulative definition, a notational convenience which, paralleling modern notations, might have been better rendered as

$$\langle z, w\rangle \in \hat{x}\hat{y}(\varphi!xy) = df\ \varphi!zw.$$

On a realist semantics for *Principia*, $\varphi!xy$ would depict a relation in intention standing between x and y. In contrast, zRw indicates that R is a relation in extension expression to be contextually defined. These emulate relations in extension of individuals. (They do not emulate relations-in-extension of classes or of relations in extension, and for that *Principia* imagines further definitions.)

Whitehead obviously approved of Russell's no-classes and no-relations-in-extension theories in *Principia*. But somehow he later seems to have changed his mind! This might have been in part a reaction to an experimental system that obviated *Principia's* no-class theory which Russell put the new introduction to *Principia's* second edition. In the experimental system, a new grammar is adopted and extensionality is an axiom schema. In 1934, Whitehead came to believe that *Principia's* no-relation in extension theory is inadequate. It fails to capture relational order. He published a paper attempting to rectify the situation and it appeared as: "Indication, Classes, Numbers, Validation," *Mind* 43, (1934), pp. 281–297. Whitehead wrote:

> The use of the symbol $\hat{x}\hat{y}\,\varphi(x, y)$ in *Principia Mathematica* involves the presupposition that the linear space-order involved in $\hat{x}\hat{y}$ can *assign* (as distinct from *symbolize*) an order to the specific functions of x and y in $\varphi(x, y)$. But the definition of linear space-order in logical terms has not, at this stage of exposition, been effected. Thus $\hat{x}\hat{y}\varphi(x, y)$ is infected with the intension involved in visual experience, in respect to its meaning. This criticism was explicitly formulated by Prof. H.M. Sheffer in his review of *Principia Mathematica*, vol. 1, 2nd edition, in *Isis*, vol. viii (I), February, 1926.

It is quite interesting to wonder what on earth Whitehead had in mind and how he might have imagined a replacement for *Principia's* treatment for its expressions for relations in extension.

One thing is certain. He was quite mistaken. In a wonderful letter to Russell, Quine had occasion to mention Whitehead's mistake. Quine had sent Russell his book, *A system of Logistic* (1935), and Russell had returned a letter of reply congratulating him. Quine replied:

> I disagree with Sheffer's claim that the spatial order of symbols smuggles a primitive idea of "order" into the "*xRy*" and "*yRx*" of PM, and likewise I hold that in distinguishing *x,y*, from *y,x* I have recourse to no such concealed primitive. …Sheffer's stand is an example, to my mind, of what might be termed the "introspective fallacy"—the same fallacy, e.g., whereby students erroneously object to Sheffer's own stroke-function that it involves two primitive, "either—or—" and "not", in view of the verbal explanation of the stroke as "either not—or not.—" In either case the fallacy depends upon venturing too far from the pole of the formalist, who asks, regarding primitives, only: "What notational devices are not introduced by abbreviative conventions in terms of previous notational devices?" (Quine to Russell 4 July 1935; quoted from Grattan-Guinness 2000, p. 591).

Quine also addressed this issue on different occasion, reflecting upon the incident:

> Russell expressed misgivings also over my abandonment of propositional functions in favor of classes and my elimination of his axiom reducibility. Also he echoed, tentatively, an allegation of Sheffer's that there is a circularity in the notion of ordered pair. These three misgivings are all traceable to a failure to maintain a sharp distinction between use and mention of expressions—a failure that had likewise caused the fogginess and complexity of early portions of *Principia Mathematic*. It was not to be wondered that my answer to Russell's 600-word letter was three times that long." (From *Quine in Dialogue*, ed., by Douglas B,. Quine and Dagfinn Føllesdal (Harvard University Press, Cambridge, 2008), p. 100.)

While there is no reason to agree with Quine that *Principia's* informalities about use and mention were founded in any fogginess on Russell's part, it is quite wonderful to realize that Quine took a strident stand against Whitehead's concern that there is some problem of capturing order in *Principia's* no-relations in extension theory.

It is worth noting that the Wiener-Kuratowski constructions (1912/1914) of ordered pairs is of no importance for *Principia's* mathematical logic. The idea of the Wiener-Kuratowski's construction can be represented by the following stipulative definition:

$$\langle z, w\rangle = df\{u: u = \{z\} \vee u = \{z, w\}\}.$$

As we can see, if this were couched within simple type theory, it would undermine the construction of expressions for non-homogeneous relations in

extension. Non-homogeneous relations are of central importance in *Principia*. They enable the cardinalities of classes of different relative types to be compared. For example, the class of natural numbers

$$\{0, 1, 2, \ldots n, \ldots\}$$

has the *same* cardinality as the class of singletons of natural numbers

$$\{\{0\}, \{1\}, \{2\}, \ldots\{n\}, \ldots\}.$$

The comparison requires the existence of a non-homogeneous relation R which can have relata such that $0\ R\ \{0\}$ and $1\ R\ \{1\}$ and so forth. Indeed, non-homogeneous relations give rise to the non-homogeneous relation of *sm* (similarity) which is such that

$$\{0, 1, 2, \ldots n, \ldots\}\ sm\ \{\{0\}, \{1\}, \{2\}, \ldots\{n\}, \ldots\}.$$

This says that the class of natural numbers has the same cardinality as the class of singletons of natural numbers. The viability of such comparisons of the cardinalities of classes of different relative types is one of the gems of *Principia*. It would be entirely lost if the techniques of Wiener-Kuratowski were adopted.

I would hope that Whitehead fully understood this. His break with Russell over the no-relations in extension theory marks a very quirky episode of intellectual disagreement between Russell and Whitehead. Indeed, in volume II of *Principia*, Whitehead worked at great length to develop a theory of non-homogeneous *similarity* relations. As we shall see, one of Whitehead's discoveries about cardinals is that when they are couched in *Principia's* theory of non-homogeneous relations, cardinals do *not* always obey Hume's Principle. The principle, so dear to neo-Fregeans, is this:

$$\mathrm{Nc}`\alpha = \mathrm{Nc}`\beta \equiv \alpha\ sm\ \beta.$$

This says that the cardinal number of α is identical to the cardinal number of β if and only if α is similar to β.

The reason it is not always true is that, where non-homogeneous relations of similarity are involved, Cantor's power-class theorem sometimes prevents it. This occurs because Cantor's power-class theorem enables a proof in *Principia* that some descending cardinals are empty.

Division of Labor in *Principia*

Whitehead had been a staunch advocate of making the notations of *Principia* as close as possible to the notations of classes that were popular among working mathematicians of the time. The working mathematicians are not detained by

paradoxes of classes because the notations were merely a tool for the investigation of the mathematical structures under study. *Principia's* no-classes and no-relations theories were a great relief to Whitehead. Indeed, as early as 1905 Whitehead had reminded Russell of this:

> This extreme rigour must be tempered by practical considerations. Classes can be kept in use by the consideration that the object is to systematize the actual reasoning in mathematics, and this actual reason does in fact habitually employ classes when it need not do so. Thus our object is to systematize the reasoning concerning classes, even when it is a primitive idea which might be avoided.

Principia's class notations were something of a compromise between the two authors over foundations and give a nice portrait of their collaborative efforts in the field.

Russell was more than happy to acknowledge Whitehead's life-long mentorship and their collaboration on *Principia*. He wrote:

> I take this opportunity to protest Mr. Keynes's practice of alluding to *Principia Mathematica* as though I were its sole author. Dr. Whitehead had an equal share in the work, and there is hardly a page in the three volumes which can be attributed to either of us singly.[9]

But the details are much more interesting. In a somewhat more careful exposition of the relative contributions to the work, Russell remarked (1948 *Mind*):

> In the early parts of *Principia*, Whitehead contributed the treatment of apparent variables and the notation (x). φx. Chapters 10, 11, 13 of *Principia* are in the main his work. He also invented the notations D'R, R'x, R "α – in this last case, the concept as well as the notation. (My previous attempts at relational notation, which were clumsy, will be found in Peano's *Review de Mathématiques*, Volumes VII and VIII.)
>
> In the later parts, the primary responsibility was Whitehead's as regards cardinal arithmetic and mine as regards relation arithmetic. Whitehead alone was responsible for the section on Convergence and Limits of Functions, and for Part VI, on Quantity. Whitehead also contributed some portions which might have been thought to be more in my province, for instance the "Preparatory Statement of Symbolic Conventions" at the beginning of Volume II, which concerned with types and systematic ambiguity. He also wrote the bulk of the first chapter of the Introduction.

In what follows we shall take these paragraphs in turn. For the present, let us consider the first.

Whitehead may well have been responsible for the many odd passages in sections *10 and *13 (and even in section *9) that discuss *Principia's conventions* for reading notations that are ambiguous concerning types of individuals. But it is worth pointing out that Whitehead was surely not responsible for the quantification theory of *Principia's* section *10. A quantification theory akin to that of section *10 is set out in Russell's 1905 "The Theory of Implication" (published in 1906). And Russell was surely responsible for the quantification theory of *Principia's* section *9's by means of which the system of *10 was to be a mere technical convenience.

[9]Russell (1922).

The system of *9 appears in Russell's work notes of 1906 (such as "The Paradox of the Liar") and it played a central role in the 1906 substitutional theory of propositional structure Russell set out in "On 'Insolubilia' and Their Solution by Symbolic Logic." Indeed, it is very likely that Russell invented the notation $(x).\varphi x$, departing from the demand that the bound variable always be appended to the horseshoe as in Peano's notation of formal implication "$\varphi x \supset_x \psi x$." It appears in letters to Frege of 1904 and also in Russell's (1903b) letters to and from Couturat.[10] Perhaps Russell meant to say that Whitehead came up with $(\exists x).\varphi x$.

It is difficult to imagine that it was Whitehead who invented the notation $R`x$, since this use of the apostrophe is used by Russell to distinguish his theory that there are no functions in intension (and no functions in extension) from Frege's formal systems that take the function sign "fx" as primitive and accept an ontology of functions in extension. For Russell, some relations in intension are functional, and thereby so are some relations in extension. For example, where R is a two-place relation in extension, Russell has:

$$^{*}71.17 \vdash R \in 1 \rightarrow \mathrm{CIS} \equiv (x, y, z)(xRy \;\&\; xRz .\supset. y = z).$$

Russell exploits his theory of definite descriptions, writing

$$R`y = df\,(\iota x)(xRy).$$

The notation D`R occurs in Russell's 1907 paper "Mathematical Logic as Based on the Theory of Types."[11] It is an application of the distinctly Russellian idea of respecting the fact that "the domain of the relation R in extension" is a definite description. *Principia* has the following:

$$\begin{aligned} \mathrm{D} &= df\,\hat{\alpha}\widehat{R}(\exists y)(\alpha = \hat{x}(xRy)) \\ \alpha DR &= df\ \alpha = \hat{z}(\exists x)(zRx) \\ D`R &= df\ \ (\iota\alpha)(\alpha\,\mathrm{D}\,R) \end{aligned}$$

Thus, in *Principia*, the theorem

$$\vdash \mathrm{D}`R = \hat{z}\,(\exists x)(\mathrm{z}Rx)$$

is substantive and employs both the 1905 theory of definite descriptions as well as the no-classes theory. The notation "$D`R$" is surely not due to Whitehead. Perhaps, Russell meant to speak of Whitehead inventing the notation $\mathrm{C}`R$ which uses the backward "D" for the range of R.

There is a curious tension in *Principia* as Whitehead and Russell sparred over the use (or what Whitehead must have regarded as the overly formal use) of definite

[10] See Schmidt (2001, p. 325).

[11] See Russell (1908).

descriptions. Whitehead clearly preferred the more convenient practice among mathematicians of putting

$$Domain\,(R) = df\ \hat{x}(\exists y)(xRy).$$

In the main, he conceded the point to Russell. But when it comes to the definitions of *sum* and *product*, there is a sudden departure in *Principia* from the use of definite descriptions. In *Principia*, the definitions of *sum* and *product* are given (respectively) as follows:

$$s`\kappa = df\ \hat{x}(\exists\alpha)(\alpha \in \kappa\,\&\,x\ \in \alpha)$$
$$p`\kappa = df\ \hat{x}(\alpha)(\alpha \in \kappa \supset x\ \in \alpha).$$

More formally they should have been defined with definite descriptions. We find Whitehead (or perhaps it was Russell) offering an apology (Whitehead and Russell 1957, vol. I, p. 303):

> Instead of defining $p`\kappa$, $s`\kappa$ ... it would be formally more correct to defined p, s, …. where the relations giving rise to the above descriptive functions. Thus we should have
>
> $$p = \hat{\beta}\hat{\kappa}(\beta = \hat{x}(\alpha \in \kappa \supset_\alpha x \in \alpha))\mathrm{Df}.$$
>
> whence we should proceed to
>
> $$\vdash \beta\, p\, \kappa \equiv \beta = \hat{x}(\alpha \in \kappa \supset_\alpha x \in \alpha)$$
> $$\vdash p`\kappa = \hat{x}(\alpha \in \kappa \supset_\alpha x \in \alpha)$$
> $$\vdash \mathrm{E}!p`\kappa.$$

This tension in the approach is interesting. Whitehead seems to have sometimes forgotten which were which.

In Volume II, when it came to defining D ‘N_0c, Whitehead forgot that one has to first define the relation sign N_0c. In *Principia*, the notion of cardinality, characteristically ambiguous in relative type, is defined with the following:

$$\mathrm{Nc} = df_{*100.01}\,\overrightarrow{sm}$$
$$\mathrm{Nc} = df_{*100.02}\mathrm{D}`\mathrm{Nc}.$$

These are fine as they stand. NC is the domain of the relation Nc. The notation D ‘R, requires that R be a relation sign. The cardinal number of α is expressed as Nc ‘α, which is $\overrightarrow{sm}$ ‘α. This leaves undecided whether the relation *sm* of similarity is ascending, homogeneous, or descending. To introduce notations for these, Whitehead offers N^{i}c and N_{o}c and N_{i}c, respectively. None of these are properly defined. If ρ in $\hat{\rho}$ (ρ *sm* α) is i-many relative types above the relative type of α, then we have the ascending cardinal N^{i}c. If, on the other hand, ρ in $\hat{\rho}$ (ρ *sm* α) is the same relative type as that of α, then we have the homogeneous cardinal N_{o}c ‘α. However,

if ρ in $\hat{\rho}$ (ρ *sm* α) is *i*-many relative types lower than the relative type of α, then we have the descending cardinal N_ic.

For example, in volume II, in the section on homogenous cardinals, one expects to find definitions for homogeneous cardinals N_oc and N_oC that are analogs of definitions *100.1 and *100.02. I imagine they were in the original drafts for volume II before Whitehead made his emendations. Unfortunately, Whitehead's emendations to volume II do not render the proper analogs. We find the following instead:

$$N_oc\,`\alpha = df_{*103.01} Nc\,`\alpha \cap t\,`\alpha$$
$$N_oc = df_{*103.02} D\,`N_oc.$$

The first does not define a relation N_oc. Thus the second remains undefined because the relation N_oc is never defined in *Principia*. Whitehead seems to have forgotten how the domain is defined in *Principia*, confusedly thinking that he has

$$NC = df\,\hat{\sigma}(\exists\alpha)(\sigma = Nc\,`\alpha).$$
$$N_oC = df\,\hat{\sigma}(\exists\alpha)(\sigma = N_oc\,`\alpha).$$

See Whitehead (Whitehead and Russell 1957, vol. II, p. 5). This mistake is only the tip of an iceberg of bad amendments.

Whitehead's Emendations of *Principia's* Volume II

Now the notations N^ic and N_oc and N_ic do not permit the designation of the sameness of relative type but rather give a numeral expressing distance (above or below) the relative type of α. In order to enable a determination of relative type, Whitehead writes $Nc(\xi)\,`\alpha$. This tells us that $\rho$ in $\hat{\rho}$ ($\rho$ *sm* $\alpha$) has the same relative type as $\xi$. Notations for relative type are essential to stabilizing unruly expressions. The notation $Nc`\alpha$ is a definite description

$$Nc`\alpha = \mathrm{df}(\iota\sigma)\left(\sigma\,\overrightarrow{sm}\,\alpha\right).$$

It is horrifically unruly in its being indeterminate in its relative type because it can change its relative type designation in different occurrences in the same proposition, and can change its relative type in different lines of proof. In contrast, Whitehead invented the notation $Nc(\xi)`\alpha$. This is also a definite description, namely:

$$Nc(\xi)`\alpha = \mathrm{df}(\iota\sigma)\left(\sigma\,\overrightarrow{sm_\xi}\alpha\right).$$

This definite description is for the class of all classes of the same relative type as ξ that are similar to α. The notation $\mathrm{Nc}(\xi)`\alpha$ is determinate in relative type. Its occurrences are uniform in the lines of a proof and in a given propositions. For example, the notations of relative type enable Whitehead to prove:

$$\xi \in \mathrm{Nc}(\xi)`\alpha \equiv \xi\, sm_{\xi}\, \alpha.$$

But one cannot prove, and one doesn't want to have:

$$\xi \in \mathrm{Nc}\,`\alpha \equiv \xi sm\, \alpha.$$

The difference is of utmost importance. Whitehead added to volume II a *Prefatory Statement of Symbolic Conventions* enabling him, in many cases, to use $\mathrm{Nc}`\alpha$ instead of $\mathrm{Nc}(\xi)`\alpha$, and leave to his conventions to determine which occurrences of $\mathrm{Nc}`\alpha$ are uniform in a proof or within a given proposition.

Relative type notations are not simple-type notations. The notations for relative types were undoubtedly due to Whitehead. They were designed specifically for the discussion of Cardinals in *Principia's* volume II. Unfortunately, they are set out in *Principia's* volume 1, sections *63–*65 hundreds of pages before they are used in volume II.[12] Their appearance in volume 1 has unfortunately tended to cause confusions—buttressing objections to type theory, with some interpreters imagining that Whitehead and Russell allow, contrary to explicit statements otherwise, that *Principia* embraces a property '… is a type' and a relation of '…is the same type as…'.[13] This is far from the case. *Principia* is a no classes theory. Accordingly, the use of lower case Greek α, β etc., and $\hat{\rho}\psi\rho$ for classes of some indeterminate relative type never have type indices restored to them. Only individual variables and predicate variables have type indices restored to them. Expressions for classes, relations in extension, definite description expressions and lower case Greek, do not have type indices at all.

With notions for relative type in place, Whitehead came upon a startling discovery about relative types: namely, that Russell's cardinals do *not* always obey Hume's Principle. The principle, so dear to neo-Fregeans, is this:

$$\mathrm{Nc}`\alpha = \mathrm{Nc}`\beta \equiv \alpha\, sm\, \beta.$$

This says that the cardinal number of α is identical to the cardinal number of β if and only if α is similar to β. *Principia* readily proves the right-to left direction:

$$^{*}100.321 \vdash \alpha\, sm\, \beta \supset \mathrm{Nc}\,`\alpha = \mathrm{Nc}`\beta$$

[12]A relative type notation does occur in volume 1 at the opening *52 and in theorem *84.21. But these are could have readily been omitted.

[13]Hylton (1990, p. 317).

i.e.,

$$\vdash \alpha \, sm \, \beta \supset \mathrm{Nc}(\breve{\xi})\,\text{`}\alpha = \mathrm{Nc}(\breve{\xi})\,\text{`}\beta.$$

But Whitehead goes on to write:

$$\text{Note that } \mathrm{Nc}\text{`}\alpha = \mathrm{Nc}\text{`}\beta \supset \text{ is not always true.}$$

The reason it is not always true is that Cantor's power-class theorem prevents it in some cases. This occurs because Cantor's power-class theorem enables a proof in *Principia* that some descending cardinals are empty.[14] Cantor's power-class theorem assures that no class α is similar to its power class (the class of all subclasses of α). This applies to V, the universal class (of a given type). That is, we have Nc(V) 'Cl 'V = Λ. It follows that we may have $\mathrm{Nc}\text{`}\alpha = \Lambda = \mathrm{Nc}\text{`}\beta$ and at the same time $\sim(\alpha \, sm \, \beta)$. For example, Nc(V) 'Cl 'V = Λ = Nc(V) 'Cl 'Cl'V. And yet we have $\sim$(Cl 'V *sm* Cl'Cl 'V).

This concern, however, was not responsible for Whitehead's halting the printing of volume II for almost a year until he could get his ideas for emendations and conventions in order. The reason, I suspect, was the difficulties Whitehead felt would accompany adhering to the consequences of his definitions of cardinal *addition*, *multiplication*, *exponentiation*, *subtraction* and *greater than*. Consider cardinal *addition*. The definition is this:

$$\mu +_{\mathrm{c}} \nu = df_{*110.02}\, \hat{\rho}(\exists \alpha, \beta)(\mu = \mathrm{N_o c}\text{`}\alpha \,\&\, \nu = \mathrm{N_o c}\text{`}\beta \,.\&.\, \rho \, sm \, (\alpha + \beta))$$

We need not go into details of the many subordinate definitions involved in this. Our point is only that among Whitehead's bad amendments we find that he introduced the following supplementary definitions for cardinal *addition*:

$$\mathrm{Nc}\text{`}\alpha +_{\mathrm{c}} \nu = df_{*110.03} \mathrm{N_o c}\text{`}\alpha +_{\mathrm{c}} \nu$$
$$\mu +_{\mathrm{c}} \mathrm{Nc}\text{`}\alpha = df_{*110.03}\, \mu +_{\mathrm{c}} \mathrm{N_o c}\text{`}\alpha.$$

The supplementary definitions are illicit. They introduce multiple d*efiniens* for the same *definiendum*. The notation $\mathrm{Nc}\text{`}\alpha$ is a definite description $(\iota\sigma)\,(\sigma\ \overrightarrow{sm}\ \alpha)$ and as such its occurrences already are given contextual definitions. For example, we have:

$$\mathrm{Nc}\text{`}\alpha +_{\mathrm{c}} \nu = df\, (\iota\mu)(\mu \mathrm{Nc}\text{`}\alpha) +_{\mathrm{c}} \nu.$$

Whitehead's introduction of illict supplementary definitions is repeated in volume II for cardinal *multiplication*, cardinal *exponentiation*, cardinal *subtraction* and for the *greater-than* relation. Similar illicit definitions are also adopted for ordinal numbers as well and even enter into *Principia's* volume III.

[14]There is no proof in *Prinicpia*, however, hat some finite descending cardinals are empty.

Many of the shocking features of Whitehead's amendments went unnoticed until Grattan-Guinness pointed them out.[15] It is a near disaster. But it can be easily rectified by simply dropping Whitehead's illicit supplementary definitions and living without them. Of course, this means living without *Principia's* theorem:

$$^{*}110.3 \vdash \mathrm{Nc}\,`\alpha +_{\mathrm{c}} \mathrm{Nc}\,`\beta = \mathrm{Nc}\,`(\alpha + \beta).$$

But we shall have instead the following:

$$\textit{homogeneous}^{*}110.3 \vdash \mathrm{N_o}c\,`\alpha +_{\mathrm{c}} \mathrm{N_o}c\,`\beta = \mathrm{Nc}\,`(\alpha + \beta).$$
$$\textit{ascending}^{*}110.3 \vdash \mathrm{N^m}\mathrm{c}\,`\alpha +_{\mathrm{c}} \mathrm{N^m}\mathrm{c}\,`\beta = \mathrm{Nc}\,`(\alpha + \beta).$$

Perhaps we can see why, in composing the second edition of *Principia*, Russell ultimately agreed with Whitehead's plea: "I don't think Types are quite right… But for heaven's sake don't alter them in the text."

At times, Whitehead seems to worry that a kind of unruly ambiguity, akin to that of "Nc" and "*sm*," might even infest *Principia's* use of lower-case Greek α, β, etc., and its use of expressions $\hat{\rho}\psi\rho$ for classes of classes (of some relative type). Of course, if uniformity of relative type is not guaranteed for lower-case Greek α, β, etc., and expressions $\hat{\rho}\psi\rho$, then deduction in *Principia* is wholly undermined. I am happy to report, however, that Whitehead's worry was misguided. Multiple occurrences of lower-case Greek in proofs and in propositions are completely *uniform* in their ambiguity. Each occurrence must is assigned the very same relative type. The uniformity derives from the uniformity of the of use schematic letters φ, ψ, *f, g*, etc. in *Principia*. They must stand in for *wffs* of *Principia* that are determinate in relative type. Thus, for example, *Principia's* theorem schema

$$\alpha \in \hat{\rho}\psi\rho \equiv \psi\alpha$$

does not yield

$$\alpha \in \hat{\rho}(\rho\, sm\, \alpha) \equiv \alpha\, sm\, \alpha,$$

The schematic use of ψ requires determination of relative type and thus a proper instance is:

$$\alpha \in \hat{\rho}\left(\rho\, sm_{(\xi,\alpha)}\, \alpha\right) \equiv \alpha\, sm_{(\xi,\alpha)}\, \alpha.$$

All is well. But it all depends essentially on *Principia* notions of relative type and its conventions for suppressing and restoring relative types.

The conventions for suppressing and restoring relative types are, in fact, essential to the intelligibility of deduction in *Principia* with Nc'α. Finding the proper conventions, however, is no small task. And Whitehead's emendations for

[15]Grattan-Guinness (2000, p. 584).

Principia's volume II were unfortunately guided his belief that if a proposition is "true whenever significant" it ought to be provable. In particular, he was misled by "$\alpha \in \mathrm{Nc}\ `\alpha$" which is certainly true whenever significant. Whitehead accepts

$$^{*}100.3 \vdash \alpha \in \mathrm{Nc}`\alpha.$$

This gets Whitehead into hot water. Struggling to keep straight what conventions ought to be adopted for suppressing and restoring relative type determinations, Whitehead finds himself proclaiming that it is a fallacy to infer $\vdash \exists!\mathrm{Nc}`\alpha$ from $\vdash \alpha \in \mathrm{Nc}`\alpha$. Further Whitehead goes on to warn readers (Whitehead and Russell, vol. II, p. 34):

> Where typically ambiguous symbols occur in implications… the conditions of significance may be different for the hypothesis and the conclusion, so that fallacies may arise from the use of such implications in inference. *E.g*., it is a fallacy to infer "$\vdash \exists!\mathrm{Nc}\ `\alpha$" from the (true) propositions "$\vdash \alpha \in \mathrm{Nc}\ `\alpha \supset \exists!\mathrm{Nc}\ `\alpha$" and "$\vdash \alpha \in \mathrm{Nc}\ `\alpha$".

In short, in his emendation of volume II of *Principia* Whitehead finds himself is the odd situation of seeming to be denying an instance of the impeccable inference rule *modus ponens*. This was a mistake. Whitehead should not have accepted $\alpha \in \mathrm{Nc}`\alpha$, since it means $\alpha \in \mathrm{Nc}(\xi)`\alpha$. It is true whenever significant but not provable. In contrast, the following are provable

$$*103.12 \vdash \alpha \in \mathrm{N_o c}\ `\alpha.$$

$$\textit{homogeneous} * 103.12 \vdash \alpha \in \mathrm{Nc}(\alpha)\ `\alpha.$$

Moreover, the following is also provable

$$\vdash \alpha \in \mathrm{Nc}(\xi)\ `\alpha \supset \exists!\mathrm{Nc}(\xi)\ `\alpha.$$

But these are obviously not of the proper form to generate an inference by Modus Ponens.

Principia's 1925 Second-Edition

Early in 1920 Russell came to realize that *Principia's* volume I was out of print. One might have hoped for the second edition of *Principia* to correct some of Whitehead's bad emendations of volume II. But nothing of the sort happened. Indeed, quite to the contrary, on 21 October 1923 Russell wrote to S.C. Roberts at Cambridge University Press as follows:

> I think it will be quite possible to proceed with the composition of Vol. II as soon as that of Vol. I is completed (except the Introduction). In addition to the Introduction, we wish to add a few notes on single chapters. If you like, these can come immediately after the

> introduction, thought they would come more naturally at the end of the volume. But they cannot well be ready sooner than the introduction.
>
> While I am in America (Jan-April) it will be necessary to get Dr. Whitehead to attend to the proofs. He also has ideas about the prefatory matter, but I think he and I can arrange that without troubling you. The corrections for Vol. II will certainly be very slight, in fact I doubt if there need be any.

The odd circumstance of Whitehead's not being involved with the 1925 the second-edition of *Principia* is recounted by Lowe who notes that, although he was overloaded with academic duties at the University of London, Whitehead did try to collaborate in some measure.

It clear, however, that something put an end to their having any significant collaboration on the second edition. Indeed, Whitehead eventually became irritated enough to write a note to the editor of *Mind* clarifying the origins of the new material for the second edition. Whitehead wrote:

> The great labour of supervising the second edition of the *Principia Mathematica* has been solely undertaken by Mr. Bertrand Russell. All the new material in that edition is due to him, unless it shall be otherwise expressly stated. It is also convenient to take this opportunity of stating that the portions in the first edition- also reprinted in the second edition—which correspond to this new matter were due to Mr. Russell, my own share in those parts being confined to discussion and final concurrence. The only minor exception is in respect to *10, which preceded the corresponding articles. I had been under the impression that a general statement to this effect was to appear in the first volume of the second edition.[16]

Lowe explains any other letters and Whitehead's notes for the second edition seem not to have survived the general destruction of Whitehead's papers after his death. So we are left to speculate upon what intellectual disagreement may have been driving things.

There is a long-standing myth to the effect that Russell's introduction to the second edition and its new appendices show him with the intent to abandon the theories of the first edition in favor of the *Tractarian* doctrines of Wittgenstein. This is far from the case. Russell explicitly notes in the introduction to the second edition of *Principia* that his investigation of those ideas, in particular the idea that a *function can occur only through its values*, concluded with the result that they fail—they are inadequate to recover Cantor's work and Analysis in general. (They do recover, in Russell's judgment, mathematical induction and the theory of natural numbers.)[17]

Russell's assessment upset Wittgenstein. Not surprisingly, Wittgenstein felt that his ideas were not properly represented in Russell's investigation and admittedly, Russell wholly ignored (because he rejected) Wittgenstein's thesis that identity is a pseudo-predicate—a thesis that would, Wittgenstein thought, require that *Principia* be done "afresh" so that arithmetic consists, not of logical truths, but of *equations* showing the sameness of calculations of outcomes of (recursive) functions.

---

[16]Whitehead (1926b).

[17]For a detailed discussion of the second edition, see Landini (2013).

In a letter to his mother of 1923, Ramsey makes this clear, writing that he had reported his knowledge of Russell's work notes for the second edition of *Principia* to Wittgenstein and he writes that "he [Wittgenstein] is, I can see, a little annoyed that Russell is doing a new edit[ition] of *Principia* because he thought he had shown Russell that it was so wrong that a new edition would be futile. It must be done altogether afresh."[18] Wittgenstein thought that his elimination of identity was his great achievement, and it is precisely that elimination that requires arithmetic to be understood in terms of equations, not tautologies. Ramsey further corroborates this in a letter to Wittgenstein of 20 February 1924:

> *I went to see Russell a few weeks ago*, and am reading the manuscript of the new stuff he is putting into *Principia*. You *are right that it is of no importance; all that it really amounts to a clever proof of mathematical induction without using the axiom of reducibility*. There are no fundamental changes; identity is just as it used to be. … of all your work, he seems now to accept only this: that it is nonsense to put an adjective where a substantive ought to be which helps in his theory of types.[19]

Curiously, Wittgenstein came to think Ramsey a traitor as well. Ramsey came to agree with Russell that Wittgenstein's stand on identity and his thesis that arithmetic consists of equations is hopeless. In his paper, "The Foundations of Mathematics," he wrote that he had spent a lot of time developing Wittgenstein's construal of identity and the theory of arithmetic as equations and found it to be "faced with insuperable difficulties." According to Ramsey, a compromise on identity must be found, and arithmetic consists of tautologies after all. Ultimately, Ramsey offered a new infinitary semantics for *Principia* and concludes: "By using these variables [function in extension], we obtain the system of *Principia Mathematica* simplified by the omission of the axiom of Reducibility and a few corresponding alterations. Formally it is almost unaltered; but its meaning as been considerably changed."[20] *Principia* wouldn't have to be done afresh after all.

Lowe offers an explanation of why Whitehead seems to have withdrawn his collaboration on the second edition of *Principia* He writes: "It is hard to imagine any greater contrast between philosophers than that between the author of *Science and the Modern World* and the author of the *Tractatus*. If Russell did not sense this, it was because Wittgenstein was the object of his current enthusiasm, while Whitehead, Russell thought, had gone in for bad metaphysics." Lowe strongly suggests that Whitehead was upset with Russell for endorsing Wittgenstein's views in the second edition. This cannot be correct. The fact on the ground is that Russell found Wittgenstein's ideas to be a dead ends and certainly did not endorse them in the second edition of *Principia*. Lowe may be correct, however, that Whitehead would naturally find it objectionable for Russell to put into *Principia*'s second edition an evaluation of his experiments with Wittgenstein.

[18]Ramsey (1923, p. 78).
[19]Ramsey (1924, p. 84).
[20]Ramsey (1931, p. 17).

Relativity, Mind and Matter

It must have been disappointing for Russell that while he was working on a second edition for *Principia*, Whitehead still hadn't finished volume IV on Geometry. Whitehead maintained that he had done a good deal of work on volume IV in England and intended to finish it when he was appointed at Harvard in 1924. Lowe reports that as late as 1930, Whitehead still imagined that he would complete the fourth volume. Precisely why the fourth volume was never completed remains something of a mystery.

Lowe is quite convinced that Whitehead did not share Russell's view that geometry was part of pure mathematics and that his grappling with the views of Einstein in the 1920s impacted the contents of what would have been volume IV. He writes (Lowe 1990, p. 94):

> … quite clearly the whole conception of Geometry as the logical analysis of space required rethinking in light of the Special Theory of Relativity. Whitehead was greatly affected by the revolution in physics that had taken place in the first decade of the twentieth century, but for him the physical conception of the interrelations of space, time and matter that emerged was far too narrow. So, I take it, nothing was more natural than to postpone the completion of the *Principia* … For the time, the more interesting and challenging question was, What are the foundations of geometry considered not as a purely mathematical, but as a physical science?

The thesis that geometry is part of *pure* mathematics is surely central to the logicism of *Principia* that Whitehead shared with Russell. It remains difficult to believe that Whitehead would have abandoned logicism in the 1920s and considered geometry a synthetic *a posteriori* science dealing with physical space. Moreover, as we shall see, nothing in Einstein's Special Theory of Relativity, which appeared in 1905, seems to have impacted what Whitehead said about Geometry in his publications and letters through 1914. We shall find that Lowe is correct that Einstein's philosophy impacted Whitehead's work on volume IV of *Principia*.[21] We must, however, find an understanding of how it impacted Whitehead's thinking about geometry as a part of *pure* mathematics.

The way to start is to realize, first and foremost, that Russell and Whitehead did not agree about the nature of matter. Indeed, one clear consequence of Whitehead's engagement with the philosophical problems of matter was that it dissuaded him from sharing his ideas (if not also collaborating) with Russell. Lowe recounts this by explaining that by 1915 "Whitehead had then been led by the work of Einstein and Minkowski to turn [away] from the composition of Volume IV of *Principia Mathematica* on geometry [the notes for which he had been sending to Russell] to that of his first book on the philosophy of physics. He now declined to send Russell his notes, explaining that he worked slowly and did not yet want his ideas set out by anyone—as had been done (with due acknowledgment) in 1914 in some parts of Russell's *Our Knowledge of the External World*."

[21]This view is also held by Harrell (1988, p. 144).

The full title of Russell's book was *Our Knowledge of the External World as a Field for Scientific Philosophy*. The full title is important because Russell's plan of the book was to illustrate his new scientific philosophy which took logic as its essence. The book uses the techniques of mathematical logic that he and Whitehead had developed in *Principia*, to show that *matter* (i.e., physical continuants obeying the laws of a physics) can be constructed as series of transient physical particulars, some with which we are acquainted ("sense-data") and others ("sensibilia") with which no one happens to be acquainted. Russell hoped for Whitehead's help, if not also his collaboration. His Preface contains the following acknowledgement:

> I have been made aware of the importance of this problem by my friend and collaborator Dr. Whitehead. To whom are due most all of the differences between the views advocated here and those suggested in *The Problems of Philosophy*.[22] I owe to him the definition of points, the suggestion for the treatment of instances and "things" and the whole conception of the world of physics as a *construction* rather than an inference. What is said on these topics here is, in fact, a rough preliminary account of the more precise results which he is giving in the fourth volume of *Principia Mathematic*a.

But the war opened a great distance between them and collaboration ceased. Whitehead wrote the following in a letter to Russell on January 8, 1917:

> I am sorry that you do not feel able to get to work except by the help of these notes—but I am sure that you must be mistaken in this, and there must be the whole of the remaining field of thought for you to get to work on—though naturally it would be easier for you to get into harness with some formed notes to go on.

Lowe writes: "In telling me of this letter, Russell naturally saw it only as an expression of Whitehead's feeling that he has been plagiarized in 1914. The letter reminds me of Russell's need for help from Whitehead, and shows Whitehead's recognition that Russell would go dancing away with Whitehead's ideas before Whitehead himself had put them in order."[23]

The issue here, however, is certainly not worries of plagiarism and I very much doubt that Russell took it this way. The issue is that Russell embraced neutral monism in 1917/8. This is a significant shift in Russell's view about matter and set him apart from Whitehead. To be sure, it is likely that Russell held something of a four-dimensionalist and Eternalist view on the philosophy of time as early as *The Principles of Mathematics* (1903a). Moreover, Russell's shift to neutral monsism surely did not bring about any new commitment to rejecting of the old materialistic conception of *cause* (where a cause brings about an event). But it made these commitments center stage. For Russell, the universe is not engaged in a process of unfolding or becoming directed by some internal principle. There is no causation in the old materialist sense where an event brings about other event causally. In his 1912 paper "On the Notion of Cause," Russell views the scientific notion of "cause" as a function which, when it takes as its argument a total event state of the universe

[22]Whitehead made several helpful comments and criticisms during Russell's writing of *The Problems of Philosophy*, but somehow the book fails to acknowledge him.

[23]Lowe (1985, p. 229).

at a given time, yields as value the configuration of the total event state of the universe at every other time, be it past, present, future.

Russell's neutral monism led him in the *Analysis of Mind* (1921) and in *Outline of Philosophy* (1927b) to embrace aspects of the philosophy set out in Watson's behaviorism. Brentano's *Principle of Intentionality* is flatly rejected, methodological solipsism is rejected more stridently than ever. Inference in general, and even our cognitive *understanding* of algebra, are now construed as if they were but habits inculcated by the behavioristic Law of Effect. *Consciousness* is almost entirely dismissed (as in behaviorism) as Russell focusses on faculties of an organism's engagement with an environmental niche. *Knowing* is largely a matter of reacting successfully to environmental stimuli.

Though Russell took a stand against Watson's behaviorist reduction of 'images' to micro dispositions to move one's larynx, he found behaviorism congenial to his *neutral monism*. Russell felt that *matter* (physical continuants obeying the new physics) and *minds* ('selves' obeying largely behavioristic laws) can be constructed as series of transient *physical* particulars. Russell's position is a neutral monism unlike that of Spinoza and James. Russell's neutral monism is part of his naturalism and physicalism. All the same, Russell was happy to embrace James's famous invective:

> I believe that 'consciousness,' when once it has evaporated to this estate of pure diaphaneity, is on the point of disappearing altogether. It is the name of a nonentity, and has no right to a place among first principles. Those who still cling to it are clinging to a mere echo, the faint rumor left behind by the disappearing philosophy.[24]

Put starkly, mind and matter in the traditional sense do not exist. Instead, Russell offers series of transient *physical* particulars. In Russell's eliminativistic neutral monism, neither mind or matter exist. Consciousness is dismissed, and some of the ideas of behaviorism become an ally.

Whitehead would have nothing of this. In in stark contrast with Russell, we find that Whitehead, interpreted James to mean only to say that consciousness is not an *object*. He argued that James was not denying that it exists, but affirming only that it is a function (Whitehead 1925, p. 199). In his book, *Religion in the Making* (1926a), Whitehead writes (Whitehead 1926a, p. 109):

> Now, according to the doctrine of this lecture, the most individual activity is a definite act of perceptivity. So matter and mind, which persist through a route of such occasions, must be relativity abstract; and the must gain their specific individualities from their respective routes. The character of a bit of matter must be something common to each occasion of its route; and analogously, the character of mind must be something common to each occasion of its route…. But that which the occasions have in common, so as to form a route of mind or a route of matter, must be derived by inheritance from the antecedent members of the route.

Whitehead maintains that mind and matter and change and causation *do* exist. He has no sympathies whatsoever with behaviorism. The order of the universe is no

[24]James (1904).

Humean accident in Whitehead's view, and there is temporal becoming. Whitehead best expresses his conception of nature as an organic *unity* when he speaks of what is the genuine religious insight that is left over when we "no longer shelter theology from science or shelter science from theology" (Whitehead 1926a, p. 79). Whitehead writes: "The religious insight is the grasp of this truth: That the order of the world, the depth of reality of the world, is the value of the world in its whole and in its parts, the beauty of the world, the zest of life, the peace of life, and the mastery of evil, are all bound together—not accidentally, but by reason of this truth: that the universe exhibits a creativity with infinite freedom, and a realm of forms with infinite possibilities; but that this creativity and these forms are together impotent to achieve actuality apart from the competed ideal harmony, which is God." (Whitehead 1926a, p. 119)

But let us step back a few years to 1911, before Russell had launched into his eliminativistic neutral monist philosophy of mind and matter. There is every reason to believe that at that time, Whitehead and Russell had been in agreement on the method of bringing Geometry (Euclidean and non-Euclidean) into their logicism as a study of relations. Whitehead's piece, "The Axioms of Geometry," was published in the eleventh edition of *Encyclopedia Britannica* for 1911 and it shows that volume IV of *Principia* was well under way. In "the Axiom of Geometry" we find Whitehead saying:

> … geometry is not a science with a determinate subject matter. It is concerned with any subject matter to which the formal axioms apply. Geometry is not peculiar in this respect. All branches of pure mathematics deal merely with types of relations. Thus the fundamental ideas of geometry (e.g., those of *points* and of *straight lines*) are not ideas of determinate entities, but of any entities for which the axioms are true. And a set of formal geometrical axioms cannot in themselves be true or false, since they are not determinate propositions, in that they do not refer to a determinate subject matter.

As expected, Whitehead goes on to define *Projective* versus *Descriptive* geometries in terms of the different axioms governing the relations collecting *points* into *straight lines*. It is important to note, however, that *Principia's* account of geometry does not replace it with an analytic geometry understood as an arithmetization program, making mathematical space about relations on numbers.[25] The content of the various kinds of geometry governed as a study of certain relations is preserved. In *Principia,* we find (Whitehead and Russell 1957, vol. II, p. 498)": "Section C, which stands outside the main developments of the book, is concerned with convergence and the limits of functions and the definition of a continuous function. Its purpose is to show how these notions can be expressed, and many of their properties established, in a much more general way than is usually done, and without assuming that the arguments or values of the functions concerned are either numerical or numerically measurable." Volume III of *Principia* set the stage for preserving measurement of magnitudes without having to introduce numbers.

[25]Gandon (2012).

Indeed, the whole of *Principia* can be seen as an attempt to eliminate from all branches of mathematics, as much as possible, dependence on numbers.

Whitehead tells us that in projective geometry, any two straight lines in a plane intersect and the straight lines are closed series which return to themselves; in descriptive geometry, two straight lines in a plane do not always intersect and a straight line is an open series without beginning or end. Whitehead notes that descriptive geometry can be conceived as the investigation of an undefined fundamental relation between three terms (points). Whitehead's remarks on metrical geometry in the article are of interest as well. He writes that Pasch took the idea of *congruence*, or metrical equality, of two portions of space (intuitively suggested by the transportation in space of perfectly rigid bodies) to be indefinable in terms of other geometrical concepts, but Lie subsequently proved that congruence is capable of definition by means of his theory of finite continuous groups—given coordinates have been introduced. The displacement of a rigid body can be modeled as a one-to-one transformation of all space into itself since the displacements of rigid bodies are modeled in terms of transformations of a congruence-group. *Metrical* geometry becomes the theory of the properties of some particular congruence-group selected for study (*Op cit*, p. 190). It was for this reason that Group theory was to have been found in *Principia's* volume IV. Whitehead goes on to show how "distance" can be defined by the Cayley-Klein projective technique. This is all pure mathematics, known *a priori* and part of the logicism of *Principia*.

Whitehead's comments in "The Axioms of Geometry" reveal that he does not find any compelling argument that absolute space of *points* is logically contradictory. However, Whitehead holds a relational view of space and it was always part of his program to find a way to avoid taking points as indefinable in geometry. This is perfectly compatible with Whitehead's orientation to geometry as a study of special sorts of relations. His orientation is salient in a letter of 1910 to Russell which offers the following:

> The beginning of Geometry is going beautifully —*500 on *Associated Symmetrical and Permutative Triadic Functions* is a picture. Then comes *502 *The Associated Relation of a Triadic Function*, and then *504 *Axioms of Permutation and Diversity*, and *505 *Axioms of Connection* (not yet in final shape).

There is a curious letter from Whitehead to Russell in 1913 in which we might see a development of Whitehead's ideas in geometry, but relations are still center stage. He writes:

> I have done a lot of writing for vol. IV, and now know much more about Geometry than formerly. Whitehead wrote a letter to Russell in October of that year with a rather promising tone. The whole [subject] depends on the discussion of the connective properties of multiple relations. This is a grand subject. It merges into the discussions of Cl 'ν where ν is a cardinal number, preferably inductive. I call such things multifolds.

Lowe remarks that the last extant letter concerning geometry in *Principia*, is from January 1914 and in it Whitehead proposed that he include ideas from his 1916 paper "La Théorie Relationniste de l'Espace" in the volume (Lowe 1990, p. 94). But even here there is nothing suggesting even a hint of the impact of

Einstein's special relativity according to which there are physically necessary relations between physical facts of metrical distance, acceleration, mass and the invariance of electromagnetic propagation. The thesis that these physical processes are interconnected is, of course, part of a physical empirical theory. How can they be of relevance to pure geometry?

In addressing this question, it is worth recalling that *Rational Dynamics* is part of pure mathematics for Russell and Whitehead. Rational dynamics is extensively discussed in Russell's *Principles* where the logical notions *matter*, *metrical congruence*, *motion* and *time*, are discussed in connection with Zeno's paradoxes. This discussion continued in *Our Knowledge of the External World*. Let us briefly recall some of Russell's ideas about rational dynamics in *Principles*. Russell writes:

> We can now attempt an abstract logical statement of what rational Dynamics requires matter to be. In the first place, time and space may be replaced by a one-dimensional and *n*-dimensional series respectively. Next, it is plain that the only relevant function of a material point is to establish a correlation between all moments of time and some points of space, and that this correlation is many-one [functional]. So soon as the correlation is given, the actual material point ceases to have any importance. Thus we may replace a material point by a many-one relation whose domain is contained in a certain three-dimensional series. To obtain a material universe, so far as kinematical considerations go, we have only to consider a class of such relations subject to the condition that the logical product of any two relations of the class is to be null. This condition insures impenetrability. If we add that the one-dimensional and the three-dimensional series are to be both continuous, and that each many-one relation is to define a continuous function, we have all the kinematical conditions for a system of material particles, generalized and expressed in terms of logical constants.

The natural objection to Russell's view can be voiced by simply asking the question: In virtue of what is a given series to be regarded as a *temporal* series? In virtue of what is a given *n*-dimensional series to be regarded a *space*? These questions apply to both Euclidean and non-Euclidean temporal-dimension series. The mathematics here is not about *space*; it is about relations ordering their fields. Indeed, the same may be said of analytic geometry which abandons lines and points and geometric figures in favor of relations and functions on numbers. Russell can only say that the study of these relations is of particular importance because they provide excellent mathematical *models* which can apply to physical space, and their applications are empirically well-corroborated. But, in fact, it was Russell's view that these mathematical ideas ought to *replace* the ordinary and often confused ideas of synthetic geometry that have us thinking of *lines* in terms of our experiences with sticks and *points* in terms of our experience with pebbles. The empirically grounded geometric notions are fraught with confusions, and Russell felt that thinking in these experiential ways serve only to introduce muddles into both mathematics and metaphysics.

There is yet more. In Russell's view, Zeno was correct that there is no *change* in the ordinary empirically informed sense, but that we are better off abandoning it. He writes (Russell 1901):

> After being refuted by Aristotle, and by every subsequent philosopher from that day to our own, these arguments were reinstated, and made the basis of a mathematical renaissance, by

> a German professor, who probably never dreamed of any connection between himself and Zeno. Weierstrass, by strictly banishing from mathematics the use of infinitesimals, has at last shown that we live in an unchanging world, and that the arrow in its flight is truly at rest. Zeno's only error lay in inferring (if he did infer) that, because there is no such thing as a state of change, therefore the world is in the same state at any one time as at any other. This is a consequence which by no means follows; and in this respect, the German mathematician is more constructive than the ingenious Greek. Weierstrass has been able, by embodying his views in mathematics, where familiarity with truth eliminates the vulgar prejudices of common sense, to invest Zeno's paradoxes with the respectable air of platitudes; and if the result is less delightful to the lover of reason than Zeno's bold defiance, it is at any rate more calculated to appease the mass of academic mankind.

Russell goes on in *Principles* to say that the notions of *change* and *motion* are logically subsequent to the notion of occupying a *place* at a *time*. He writes (Russell 1903a, p. 473).

> Motion consists merely in the occupation of different places at different times, subject to continuity… There is no transition from place to place, no consecutive moment or consecutive position, no such thing as velocity except in the sense of a relation number which is the limit of a certain set of quotients. The rejection of velocity and acceleration as physical facts (i.e., as properties belonging at each instant to a moving point, and not merely real numbers expressing limits of ratios) involves, as we shall see, some difficulties in the statement of the laws of motion; but the reform introduced by Weierstrass in the infinitesimal calculus has rendered this rejection imperative.

The upshot is that, for Russell, the ordinary, empirically informed, notions of *velocity* and *acceleration* are to be rejected and replaced by the mathematical notions.

It is natural enough, with the help of the constructions of Weierstrass, for Russell's logicism to reject Newton's attempt to use *fluxions* (which were physical dynamic processes of moving bodies) to justify the mathematics of *limits* and *infinitesimals* in the calculus. Weierstrass gave a foundation for limits without any appeal to infinitesimals or dynamic physical processes. Analysis has been set on a firm foundation by mathematical philosophers such as Dedekind, Weierstrass, and Cantor and it has been set free from the muddles derived from confusions of thinking in terms of our experience with physical objects and motions. But there is no reason to agree with Russell that this has the implications that physics should get along without embracing uniquely physical notions of *acceleration*, *mass*, *matter*, *motion* and *change*. Russell rejected the traditional position of Newton which assumes the robust *physical* reality of these notions. It remains interesting, therefore, to ponder how Russell may have imagined his views on rational dynamics to fit with Einstein's theory of Special Relativity.

Though Special Relativity appeared in 1905, we don't find Russell and Whitehead engaging with it. Quite independently of Einstein's work, Whitehead held in 1911 that while many different geometries of metrical congruence relations can be studied in pure mathematics, it is an open empirical question as to which properly models physical space. In his article "Axioms of Geometry," Whitehead flatly rejects Poincaré's thesis that the geometry of physical space is conventional. He remarks (*Op. cit.* p. 192):

> … we have, in fact, presented to our senses a definite set of transformations forming a congruence-group, resulting in a set of measure relations which are in no respect arbitrary. Accordingly, our scientific laws are to be stated relevantly to that particular congruence-group. The investigation of the type (elliptic, hyperbolic or parabolic) of this special congruence-group is a perfectly definite problem, to be decided by experiment.

The question of which geometry models physical space is an entirely empirical one for Whitehead and has a definite answer. But what meaning (if any) can be given to the notion of empty *physical* space having a metric? Russell just abandons these notions, replacing them with mathematical structures; and thus the empirical question becomes which mathematical model works best in the science observations of empirical physics. In contrast, Whitehead seems more inclined to embrace the existence of physical space and physical time.

A step in the right direction to resolving our puzzle of how Einstein's work could have impacted Whitehead's volume IV of *Principia* is to focus on his theory of General Relativity (1916). In forming his theory of General Relativity, Einstein was under the influence of Mach whose thesis of the relativity of *all* motion (inertial or accelerated) placed him in opposition to Newtonian thought experiments designed to empirically demonstrate the absoluteness of motion and acceleration.[26] One of Newton's strongest thought experiments for absolute motion was his case of a rotating bucket of water. The rotation of the bucket and water brings the water up the sides. But if motion is relative, it is perfectly legitimate to regard the bucket and water at rest and the universe as rotating around it. What force then accounts for the water's elevation up the sides? Newton believed in the reality of absolute accelerated motions. Because acceleration is absolute, there is a *privileged* coordinate system for the laws of physics. The state of motion of the coordinate system may *not* be arbitrarily chosen. If the laws of mechanics are to be valid, it must be free from rotation and acceleration. Mach maintained, in opposition to Newton, that forces due to such accelerated motions are covariant and thus acceleration is also a relative motion. The water's elevation is due to the effect of the stars rotating about the bucket.[27]

Einstein's General Relativity endeavors to make *all* motions relative, including accelerations and rotations, and it sets for the distinctly *philosophical* thesis that the form of the laws of physics must be invariant no matter the coordinate system chosen. Einstein writes: "What has nature to do with our coordinate systems and their state of motion? If it is necessary for the purpose of describing nature, to make use of a coordinate system arbitrarily introduced by us, then the choice of its state of motion ought to be subject to no restriction; the laws ought to be entirely independent of this choice (general principle of relativity)".[28] Einstein appeals, next, to pure mathematics to find such a theory. He writes:

---

[26]See Friedman (1983).

[27]Mach dismissed thought experiments that imagined the physical laws of our universe obtaining and yet with only the bucket of water. Such logically possible universes are not physical universes see Mach (1883).

[28]See Albert Einstein, "What is the Theory of Relativity," in Einstein (1934, p. 57).

> Our experience hitherto justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas. I am convinced that we can discover by means of purely mathematical constructions the concepts and the laws connecting them with each other, which furnish the key to the understanding of natural phenomena. Experience may suggest the appropriate mathematical concepts, but they most certainly cannot be deduced from it. Experience remains, of course, the sole criterion of the utility of a mathematical construction. But the creative principle resides in mathematics. In a certain sense, therefore, I hold it true that pure thought can grasp reality, as the ancients dreamed.
>
> … In order to justify this confidence, I am compelled to make use of a mathematical construction. The physical world is represented as a four-dimensional continuum. If I assume a Riemannian metric in it and ask what are the simplest laws which such a metric system can satisfy, I arrive at the relativistic theory of gravitation in empty space.[29]

This suggests a position consonant with the mathematical theory of rational dynamics suggested in Russell's *Principles*. If we bring Whitehead in line with that, we can imagine that he sees Einstein's use of the tensor calculus as a part of a *pure* mathematical theory involving transformations of a geometric metric field.

The key to understanding why Einstein's work should have impacted Whitehead's thinking for volume IV of *Principia* is to be found in Whitehead's engagement with General Relativity, not Special Relativity. Einstein's general relativity is but one mathematical system of many that Whitehead envisions, all of which (as with measurement theory) are part of pure mathematics and would be discussed in the rational dynamics that would naturally find a place in volume IV. Whitehead endeavors to cast the issues of general relativity in the realm of pure mathematics and philosophy. Whitehead writes (Whitehead 1925, p. 168):

> The new relativity associates with space and time with an intimacy not hitherto contemplated; and presupposes that their separation in concrete fact can be achieved by alternative modes of abstraction, yielding alternative meanings. The fact relevant to experiment, is the relevance of the interferometer to just one among the many alternative systems of these spatio-temporal relations which hold between natural entities. What we must now ask of philosophy is to give us an interpretation of the status in nature of space and time, so that the possibility of alternative meanings is preserved. These lectures are not suited for the elaboration of details; but there is no difficulty in pointing out where to look for the origin of the discrimination between space and time. I am presupposing the organic theory of nature, which I have outlined as a basis for a thoroughgoing objectivism. …. In organic philosophy of nature there is nothing to decide between the old hypothesis of the uniqueness of the time discrimination and the new hypothesis of its multiplicity. It is a matter for evidence drawn from observations (Cg, my *Principles of Natural Knowledge*, Sec. 52:3).

And he goes on:

> Also you will remember that the utilization of different spatio-temporal systems means the relative motion of objects. When we analyse this critical relation of a special set of events to any given event A, we find that the explanation of the critical velocity [of light in vacuo] which we require. I am suppressing all details. It is evident that exactness of statement must be introduced by the introduction of points, and lines, and instants. Also that the origin of geometry requires discussion; for example, the measurement of lengths, the straightness of

[29]See Albert Einstein, "On the Method of Theoretical Physics," in Einstein (1934, p. 18).

> lines, and the flatness of planes, and perpendicularity. I have endeavored to carry out these investigations in some earlier books, under the heading of the theory of extensive abstraction…

Both Einstein and Whitehead are appealing, in building space-time systems, to very abstract features of their respective principles of relativity. They are not appealing to physical realities of acceleration, mass, motion and change themselves. Einstein, no less than Whitehead was doing philosophy in his theory of General Relativity.[30] It is a purely mathematical rational dynamics that is involved in their each building theories of general relativity.

When the mathematics of non-Euclidean geometries appeared, the natural question arose of whether they are about physical space. The mathematics is more appropriately about relations, if not relations on numbers. But exactly the same question can be asked about the mathematics of Euclidean geometry. The question of physical space is thus left to the issue of which of these mathematical systems is most useful in predicting and explaining experiences. When it comes to the mathematics of the transformation of space through time, the situation is quite analogous. The ordinary experientially informed notion of temporal order seems quite outside of pure mathematics. And one may well say that such systems of transformations are not properly about physical space or physical time. Now if one can accept the physical law that electro-magnetic propagation in vacuo is invariant, then motions determine the relationships of both physical space and physical time. That looks very much a part of physics, not mathematics. But the idea of General Relativity steps into play and demands that one looks for a pure mathematical system governing the relativity of all motion (accelerated or otherwise). That is why Einstein's work became a legitimate study for volume IV of *Principia*.

Whitehead rejected Einstein's use of light signals to give physical *meaning* to the notion of 'simultaneity'. In his paper "Einstein's Theory" (published in 1920a), Whitehead writes:

> In view of the magnificent results which Einstein has achieved it may seem rash to doubt the validity of a premise so essential to his own line of thought. I do, however, disbelieve in this invariant property of the velocity of light, for reasons which have been partly furnished by Einstein's own later researches. The velocity of light appears in this connection owing to the fact that it occurs in Maxwell's famous equations, which express the laws governing electro-magnetic phenomena. But it is an outcome of Einstein's work that the electro-magnetic equations require modification to express the association of the gravitational and electro-magnetic fields.[31]

Nonetheless, Whitehead allows the formulation of quite different, yet purely mathematical, systems of 'simultaneity.' This is quite a novel view which, as it were, places non-Newtonian (non-absolute) time is a part of pure mathematics just as non-Euclidean Geometry became part of pure mathematics. In an undated letter to Russell, he writes about results of his current work:

---

[30]See Einstein (1961).

[31]Whitehead (1920, p. 242).

> The result is a relational theory of time, exactly on four legs with that of space. As far as I can see, it gets over all the old difficulties, and above all abolishes the instant in time, *e.g.* the present instant, even in the shape of the instantaneous group of events. This has always bothered me as much as the 'point' – but I have had to conceal my dislike from lack of hope. But I have got my knife into it at last. According to the theory, the time-relation as we generally think of it [sophisticated by philosophers] is a great look up. Simultaneously does not belong to it. That comes in from the existence of the space-relation. Accordingly the class of all points in space serves the purpose of the instant in time.[32]

In his book, *The Concept of Nature*, Whitehead writes: "This question can be formulated thus, Can alternative temporal series be found in nature? A few years ago such a suggestion would have been put aside as being fantastically impossible. It would have no bearing of the science then current, and was akin to no ideas which had ever entered into the dreams of philosophy" (Whitehead 1920a, p. 71). The question is odd. Again one must ask: In virtue of what is a given purely mathematical serial relation a *temporal* series? If a certain series is to model physical time, we shall have to make choices about what in the physical world is to be our standard, and the worry is that different legitimate choices might well be available: pendulum, vibrating crystal of piezoelectric material, electromagnetism in vacuo, and so on. In any case, it seems clear that Whitehead's thinking remains oriented to the question of what mathematical system of space-time best models physical space and time. As with physical metrical congruence of space, he remains hopeful that experiments will settle the question and he flatly rejects conventionalism about forming operational definitions for simultaneity or metrical distance. The different space-time systems available, however, are part of pure mathematics. Physics will experimentally select from among them, but as Einstein put it, "the only method of selection between them is to wait for experimental evidence respecting those effects on which the formulas differ" (Whitehead 1925, p. 173).

The foundational philosophical doctrine of Whitehead's methods for forming space-time systems is, however, the *uniform* relatedness of all of nature. This approach, like Einstein's approach, is metaphysical and not empirical. Whitehead contends that it is essential to the very concept of "nature." Whitehead regards Einstein's system of General Relativity as one among many space-time systems available within mathematics. Geometry, as Whitehead now sees it, is the science of all logically possible kinds of space. It is investigated *a priori* but grounded in the method of *extensive abstraction* (Whitehead 1919, p. 504). Physics is the science of the contingent relations in nature; it decides empirically among different metrical geometries (*PR*, p. 503).

It is important to understand that Russell was not, in principle, opposed to Whitehead's method of extensive abstraction. Indeed, he credits Whitehead with a discovery of utmost importance (Russell 1927a, p. 22):

> Ever since Greek times," he writes, "those who did not believe in the reality of 'points' were faced with the difficulty that a geometry based on points works, while no other way of

[32]Whitehead–Russell correspondence, Bertrand Russell Research Center, McMaster University, Hamilton, Ontario, Canada.

> starting geometry was known. This difficulty, as Dr. Whitehead has shown, exists no longer. It is now possible, as we shall see at a later stage, to interpret geometry and physics with material all of which is of a finite size—it is even possible to demand that none of the material shall be smaller than an assigned finite size. The fact that his hypothesis can be reconciled with mathematical continuity is a novel discovery of considerable importance…

Whitehead's new approach to defining *points* puts pure geometry on a new footing and clearly has direct relevance to *Principia's* volume IV. Moreover, it is couched within Whitehead's methods of extensive abstraction— a kind of topology where *overlapping* (of events as durations or processes) is taken as primitive. The method was set out in Whitehead's book *The Concept of Nature* (1920b) and the metrical geometries derived from it remain part of pure mathematics. Russell was influenced by these ideas, but attempted his own method of eliminating points in *The Analysis of Matter*. In "On Order in Time" (1936), Russell borrows the idea of using overlapping events to define temporal "instants."

Whitehead's commitment to the uniformity of nature (the organic view of nature as a connected system of relations among processes) is intimately tied to his methods of extensive abstraction. Russell thought this metaphysics unnecessary. The constructions of *points* and instants of *time* are given without Whitehead's metaphysical assumptions of organic relatedness. In Russell's book, *The ABC of Relativity* (1925), Whitehead's *principle of relatedness* of nature is not so much as mentioned. Whitehead set it out in *The Principle of Relativity* (1922), in "The Philosophical Aspects of the Theory of Relativity" (1922) and *Science and the Modern World* (1925). Russell's views on these topics do not emerge until *The Analysis of Matter*. Russell was less than convinced and presents it as if it were little more than a religious intuition, derived from Bergson that "scientific inference ought to be reasonable." He writes (Russell 1927a, p. 79):

> Perhaps we all make this assumption in one form or another. But for my part I should prefer to infer 'reasonableness' from success, rather than set up in advance a standard of what can be regarded as credible. I do not therefore see any ground for rejecting a variable geometry as Einstein's. But equally I see no ground for supposing that the facts necessitate it. The question is, to my mind, merely one of logical simplicity and comprehensiveness. From this point of view, I prefer the variable space in which bodies move in geodesics to a Euclidean space with a field of force. But I cannot regard the question as one concerning the facts.

But it should be said, against Russell, that Einstein did set up his own non-empirical postulate of that might equally be regarded as a postulate of "reasonableness" according to which physical laws do not require any special coordinate system (inertial frame) for their articulation. Unlike Whitehead, we find Russell accepting a form of conventionalism. The important point, however, is that both Whitehead and Russell are agreed that a new branch of mathematics is needed for General Relativity. It is, therefore, quite natural that they both would agree that volume IV of *Principia* must accommodate it.

The mathematical orientation is, therefore, the key to understanding why *Principia's* volume IV on geometry was held up while Whitehead put his thoughts in order on the impact of Einstein's General Relativity. In grappling with Einstein's General Relativity, Whitehead can naturally maintain that geometry is part of pure

mathematics and known *a priori* as a study of special kinds of relations. In the full spirit of Russell's comments in *Principles*, Whitehead may well think that General Relativity (*not* special relativity) is a part of the pure mathematics of rational dynamics. This, and not Lowe's thesis that Whitehead abandoned logicism concerning geometry, explains why volume IV of *Principia* was again and again put on hold. In 1918, Whitehead wrote to Russell:

> When I can get back to hard work— I am afraid of pushing myself just at present, and must have a week or two's rest— I shall finish a paper I am writing (for the Royal Society, I think) on Time, Space and Matter. I bring out Electromagnetic Relativity as the most natural supposition one can make about things— and the whole method, as it appears to me, throws rather a fresh light on some problems of science and of philosophy.

Because of the importance of the general theory of relativity to the philosophy of geometry, we can agree with Lowe that nothing was more natural than to postpone the completion of the *Principia* in order to, as Whitehead put it (Whitehead 1920a, p. vii): "lay the basis of a natural philosophy which is the necessary presupposition of a reorganized speculative metaphysics" (Lowe 1990, vol. 2, p. 94).

Volume IV on Geometry remains pure even accepting that General Relativity has an important impact on it. Thus, we disagree with Lowe's interpretation of Whitehead's concern. All the same, Whitehead was forming a theory of *life* and *mind* influenced by Bergson's *organic* view of nature. Russell, in stark contrast, was developing the eliminativistic positions concerning *life* and *mind* in his neutral monism influenced by Mach who imagined space-time as a manifold of 'sensations'.[33] By the time Whitehead wrote *Science and the Modern World* (1925), his ideas had become completely in opposition to Russell's. Indeed, Russell's (1926) review of the book speaks volumes about how far apart their views had become.[34] Russell characterizes Whitehead as follows:

> An organism which contains no organic parts is called a "primate." A proton, and perhaps an electron, would be an association of such primates, superposed on each other, with their frequencies and spatial dimensions so arranged as to promote the stability of the complex organism, when jolted into accelerations of locomotion.
>
> … Dr. Whitehead is profoundly influenced by Bergson's belief in interpenetration, which he even carries further, since he regards the present as containing implicitly not only the past, but the future…
>
> …Whitehead's theory consists of two parts: on the one hand, a logical construction leading to physics from a new set of non-material fundamentals, wholly admirable and profound; on the other hand, a metaphysics believed by the author to be bound up with his logical construction, but in fact—again I speak with diffidence—separable from it. The metaphysic is not essentially new; it is approximately that of Bergson or Plotinus. There is a God… No reason can be given for the nature of God, because that nature is the ground of rationality. … I cannot persuade myself that his logical reconstruction of physical concepts

[33]Eddington (1958, p. 276). See Mach (1883).

[34]Bertrand Russell, "Relativity and Religion," *Nation and Athenaeus* 39 (May 29, 1926), pp. 206–7.

> has any such tendency as he attributes to it, to restore the consolations of relation to a world desolated by mechanism….[35]

The question of Whitehead's conception of God (as an organizing principle) may be put aside, though, of course, it undoubtedly irritated Russell. The breech is not to be found there. Neither is it to be found in Whitehead's building geometry from his methods of extensive abstraction, with *points*, *lines*, *planes*, and *distance* (metrical congruence) relations and *times* defined. In various writings, we saw that Russell attempted to offer his own definitions of these notions which, he explains, were products in some degree of the influence of Whitehead's methods.

The breech is that Russell's neutral monism took an eliminativistic stance with respect to *life*, *mind, matter*, *motion*, *time*, and *change*. There simply are no such things, though the laws of the new physics and the new psychology (largely behaviorist) are recovered without them. Whitehead could not accept such a metaphysical desert landscape. Whitehead viewed nature as an organically integrated *whole* within which *life*, *mind*, and *matter* have a genuine reality. The dispute rages to this day with no conclusions reached about which has the philosophical edge over the other or whether it is Russell's temperament or Whitehead's that is the more justified.

References

Eddington, A. (1958). *The nature of the physical world*. Ann Arbor: University of Michigan.
Einstein, A. (1934). *Essays in science*. New York: Philosophical Library.
Einstein, A. (1961). *Relativity*. New York: Crown Publishers.
Friedman, M. (1983). *Foundations of space-time theories: Relativistic physics and philosophy of science*. Princeton: Princeton University Press.
Gandon, S. (2012). *Russell's unknown logicism*. UK: Palgrave-Macmillan.
Grattan-Guinness, I. (2000). *The search for mathematical roots 1870-1940*. Princeton: Princeton University Press.
Harrell, M. (1988). Extension to geometry of *Principia* mathematica and related systems II. *Russell, 8*, 140–160.
Hylton, P. (1990). *Russell, idealism and the emergence of analytic philosophy*. Oxford: Clarendon Press.
James, W. (1904). Does consciousness exist? *Journal of Philosophy, Psychology and Scientific Methods, 1*, 477–491.
Landini, G. (2013). Review of Bernard Linsky. *The evolution of Principia mathematica (Cambridge, 2011)*, *hist and Phil of logic*, *34*, 77–97.
Landini, G. (2015). Whitehead's (*Badly*) Emended *Principia. History and Philosophy of Logic*.
Lowe, V. (1985). *Alfred North Whitehead: The man and his work 1861-1910* (Vol. 1). Baltimore: The Johns Hopkins Press.
Lowe, V. (1990). In J. B. Schneewind (Ed.), *Alfred North Whitehead: The man and his work 1910-1947* (Vol. 2). Baltimore: The Johns Hopkins Press.
Mach, E. (1883). *The science of mechanics: a critical and historical account of its development.* Chicago: Open Court, 1919.

[35]Russell (1926).

Ramsey, F. (1923). Letter to Russell of 20 September 1923. in G. H. von Wright (Ed.), *Ludwig Wittgenstein, Letters to C. K. Ogden*, with an appendix of letters by Frank Plumpton Ramsey (1973, p. 78). Oxford: Basil Blackwell.

Ramsey F. (1924). Letter to Russell of 20 February 1924. in *Wittgenstein, Letters to C. K. Ogden*. Oxford: Basil Blackwell, 1973.

Ramsey, F. (1931). The foundations of mathematics. in T. Braithwaite (Ed.), *The foundations of mathematics and other essays by Frank Plumpton Ramsey (Harcourt, Brace and Company)*.

Russell, B. (1903a). *The Principles of Mathematics*. London: George Allen & Unwin, 1956.

Russell, B. (1903b). In Schmidt, *Letter to couturat* 2001.

Russell, B. (1908). Mathematical logic as based in the theory of types. *The American Journal of Mathematics, 30*, 222–262.

Russell, B. (1912). On the nature of cause. *Proceedings of the Aristotelian Society, 13*, 1–26.

Russell, B. (1914). *Our knowledge of the external world as a field for scientific method in philosophy*. Chicago: Open Court Publishing Company.

Russell, B. (1921). *The analysis of mind*. London: George Allen and Unwin.

Russell, B. (1922). Review of Keynes's treatise on probability in the mathematical gazette 11 (July, pp. 119–125). In J.G. Slater (Ed.), *The collected papers of Bertrand Russell: Essays on language, mind and matter 1919-26* (1988, pp. 113–124). London: Unwin Hyman.

Russell, B. (1925). *The A B C of relativity*. New York: Harper & Row.

Russell, B. (1926). Relativity and religion. In J. Slater (Ed.), *The collected papers of Bertrand Russell: Essays on language, mind and matter 1919-26* (Vol. 9, 1988, pp. 312–315).

Russell, B. (1927a). *The analysis of matter*. New York: Harcourt, Brace & Co. Inc.

Russell, B. (1927b). *Philosophy*. New York: W.W. Norton 7 Company, Inc. Published in Britain under the title *An Outline of Philosophy*.

Russell, B. (1948). Whitehead and *Principia mathematica*. *Mind* (Vol. 57, pp. 137–138). In J. Slater (Ed.), *The collected papers of Bertrand Russell: Last philosophical testament: 1943-1963* (Vol. 11, pp. 190–191).

Russell, B. (1952). Alfred North Whitehead. *The listener* 48 (10 July, pp. 51–52). Reprinted in J. Slater (Ed.), *The collected papers of Bertrand Russell: Last philosophical testament* (Vol. 11, 1997, pp. 191–195).

Russell, B. (2001). *Correspondance avec Louis Couturat (1897–1913)*. A.F. Schmidt (Ed.), Paris, Kimé.

Whitehead A. N., & Russell, B. (1957). *Principia mathematica* (Vol. 1, 1910, Vol. II 1912, Vol. III 1913). Cambridge: Cambridge University Press. Pagination to the 2nd edition, 1957.

Whitehead, A. N. (1911). Axioms of geometry. *Encyclopedia Britannica* (11th issue). Reprinted in *Essays in science and philosophy by Alfrerd North Whitehead* (Philosophical Library, 1948, pp. 177–194).

Whitehead, A. N. (1916). La Théorie Relationniste de l'Espace. *Revue de Metaphysique et de Morale, 23*, 423–454.

Whitehead, A. N. (1919). An enquiry concerning the principles of natural knowledge. Cambridge: Cambridge University Press. Pagination to the 2nd edition, 1925.

Whitehead, A. N. (1920a). Einstein's theory. *The Times Educational Supplement*, 12 February 1920. Reprinted in *Essays in science and philosophy by Alfrerd North Whitehead* (Philosophical Library, 1948, pp. 241–248).

Whitehead, A. N. (1920b). *The concept of nature*. Cambridge: Cambridge University Press.

Whitehead, A. N. (1925). *Science and the modern world*. New York: Macmillan Co.

Whitehead, A. N. (1926a). *Religion in the making*. New York: Macmillan Co.

Whitehead, A. N. (1926b). *Principia mathematica*: Note to the Editor of Mind. *Mind 35*, 130.

Whitehead, A. N. (1934). Indication, classes, numbers, validation. *Mind*, *43*, 281–297. Reprinted in *Essays in science and philosophy by Alfrerd North Whitehead* (Philosophical Library, 1948, pp. 227–240).

Author Biography

Gregory Landini is Professor of Philosophy, University of Iowa. He is the author of four books: *Frege's Notations*; *what they are and how they mean* (Palgrave/MacMillan 2012), *Russell* (Routledge 2011), *Wittgenstein's Apprentice with Russell* (Cambridge 2007) and *Russell's Hidden Substitutional Theory* (Oxford 1998). He has published many articles in the philosophy of logic and metaphysics. His teaching and research interests include the foundations of mathematics, the history of analytic philosophy, modal logic, philosophy of mind, and the philosophy of language.

The Place of Vagueness in Russell's Philosophical Development

James Levine

In his retrospective writings, Russell vividly describes a number of turning-points in his philosophical development. I focus here on three. First, his break with idealism:

> It was towards the end of 1898 that Moore and I rebelled against both Kant and Hegel. Moore led the way, but I followed closely in his footsteps. (Russell 1959, p. 54)

Second, his attending the International Congress of Philosophy in August 1900:

> The most important year in my intellectual life was the year 1900, and the most important event in this year was my visit to the International Congress of Philosophy in Paris. … In Paris in 1900, I was impressed by the fact that, in all discussions, Peano and his pupils had a precision which was not possessed by others. (Russell 1944, p. 12)

Third, his coming to examine the nature of symbols, specifically how symbols acquire meaning:

> During my time in prison in 1918 [for anti–war activities], I had become interested in the problems connected with meaning, which in earlier days I had completely ignored. I wrote something on these problems in *The Analysis of Mind* and in various articles written at about the same time. (Russell 1968, p. 194)

As Russell presents it, before 1918, he had been concerned not with how certain sounds we make or inscriptions we write acquire "meaning" but rather with the entities he took the meanings of certain expressions to be.

My purpose here is to examine some aspects of the relations among these three turning-points in Russell's philosophical development. In general, I believe that Russell is correct to indicate that his attending the Paris Congress in 1900 was "the most important event" in "the most important year in [his] intellectual life"—more important even than his break with idealism two years earlier. More specifically, I believe that as a result of the influence of Peano, Russell came to hold that mathematicians—in particular, Cantor, Dedekind, and Weierstrass—had solved all the traditional problems of the infinite and continuity; that in doing so and then bringing to bear mathematical techniques to philosophy, Russell's post-Peano

J. Levine (✉)
Trinity College Dublin, Dublin, Ireland
e-mail: jlevine@tcd.ie

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_7

philosophical practice undermines the Moorean philosophy he had adopted in rejecting idealism; and that in his post-1918 writings, Russell further develops and extends his post-Peano rejection of the Moorean philosophy.

This understanding of the relation between Russell's post-Idealist Moorean philosophy and his post-Peano philosophy is different from that presented by Peter Hylton in his book 1990 *Russell, Idealism and the Emergence of Analytic Philosophy*. For while Hylton distinguishes these two phases of Russell's philosophical development, he writes generally:

> [Russell's] fundamental doctrines were ones that he held before he was influenced by mathematical logic [that he acquired following the Paris Congress], and the chief effects of that influence were to enable (or force) him to articulate those doctrines further, to show him that they could play a role in the solution of problems which had previously seemed insoluble, and, especially, to enable him to defend those doctrines. (Hylton 1990, pp. 152–3)

In contrast, as I present it, far from enabling him to articulate more fully and defend the "fundamental [Moorean] doctrines" he came to accept immediately after breaking with Idealism, the technical views Russell comes to accept after the Paris Congress call into question many of those "fundamental doctrines" and play a central role in leading him to embrace the views he accepts in his post-1918 writings. I do not attempt to tell this story in full here[1]; rather, I focus more narrowly on the topic of vagueness as it develops in Russell's philosophy.

Russell's (1923) paper "Vagueness" is standardly presented as the first serious discussion of the issue within the analytic tradition.[2] He begins the paper by writing:

> Reflection on philosophical problems has convinced me that a much larger number than I used to think, or than is generally thought, are connected with the principles of symbolism, that is to say, with the relation between what means and what is meant. (Russell 1923, p. 147)

Then after writing that "[v]agueness … illustrates these remarks" (ibid.), he argues that vagueness, and its contrary, precision, are features only of representations, or symbols:

> There is a certain tendency in those who have realized that words are vague to infer that things also are vague. … This seems to me precisely a case of the fallacy of verbalism—the fallacy that consists in mistaking the properties of words for the properties of things. Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example. They have to do with the relation between a representation and that which it represents. Apart from representation, whether cognitive or mechanical, there can be no such thing as vagueness or precision; things are what they are, and there is an end to it. (Ibid., pp. 147–8)

---

[1] I attempt a fuller account (which overlaps with, but also differs in some respects from, claims I make here), in Levine (2009).

[2] See, for example, Williamson (1994, p. 37), who indicates that in Russell's (1923), "the problem of vagueness is systematically presented for the first time in something close to its current form". See also Keefe and Smith (1997, p. 1), who write that after antiquity, "there seems to have been relatively little further discussion of vagueness before Bertrand Russell's seminal paper", and include his (1923) as the first modern discussion of vagueness in their reader.

Given that Russell becomes concerned with the nature of symbols and how they symbolize only following his 1918 prison stay, and given that he regards vagueness as a phenomenon that applies only to symbols, it is natural to suppose that prior to his time in prison, vagueness does not figure in any central way in Russell's philosophy. Accordingly, in her generally sensitive and sympathetic discussion of Russell's view of vagueness, Nadine Faulkner writes:

> [V]agueness had hardly figured as philosophically important in [Russell's] earlier writings. In *The Principles of Mathematics* (1903) and *Principia Mathematica* (1910–13) there is next to nothing on vagueness. (Faulkner 2003, p. 43)

In particular, she links Russell's (1923) discussion to his post-1918 concern with symbolism:

> Ultimately, Russell's view of vagueness in 1923 is the result of a rise in the importance of symbolism in his thinking coupled with a new emphasis on psychology, (ibid., p. 45)

adding shortly thereafter:

> Russell's talk of representation, and consequently vagueness, stems in part from what in 1923 is a relatively new interest in symbolism. From 1903 to 1918, symbolism plays no significant role in Russell's philosophy. (Ibid., 47)

While it is clear that it is only in the context of his post-1918 concerns with symbolism that Russell articulates his theory of vagueness in 1923, I argue that the same notion of vagueness he presents in 1923 is central to Russell's post-Peano practice and characterization of analysis; that his post-Peano view of analysis is thereby incompatible with the Moorean conception of analysis; and that, while there are significant differences between Russell's view of analysis before and after his 1918 prison stay, in his post-Peano analyses of mathematical concepts, no less than in those after 1918, Russell is opposed to his earlier Moorean view that analysis involves, with regard to a given sentence in question, making explicit what is, in some sense, already "present to the mind" of anyone who understands that sentence, prior to analysis.

However, to make clear why it should matter how Russell's post-Peano philosophy is related to his Moorean philosophy, I begin by discussing how views of Russell's place in the history of analytic philosophy may be affected by views of his philosophical development.

Russell's Place in the History of Analytic Philosophy and His Philosophical Development

A familiar view of Russell's place in the history of analytic philosophy is that he accepted a number of views—including a foundationalist epistemology and a commitment to the "Augustinian" or "museum-myth" view of meaning, according to which the meaning of a word is an entity designated by that word, a view that

when conjoined with his "principle of acquaintance" commits him to the view that understanding a sentence requires being acquainted with the entities designated by the words in that sentence—that were severely criticized by later philosophers including Wittgenstein, Quine, and Sellars.

Thus, for example, in his 1979 book *Philosophy and the Mirror of Nature*, Richard Rorty writes:

> [T]he kind of philosophy which stems from Russell and Frege is, like classical Husserlian phenomenology, simply one more attempt to put philosophy in the position which Kant wished it to have—that of judging other areas of culture on the basis of a special knowledge of the "foundations" of these areas. (Rorty 1979, p. 8)

And the "story" Rorty "want[s] to tell" (ibid., p. 168) is how such foundationalist aspirations of Russell and Husserl were called into question by their successors:

> [I]n the end, heretical followers of Husserl (Sartre and Heidegger) and heretical followers of Russell (Sellars and Quine) raised the same sorts of questions about the possibility of apodictic truth which Hegel raised about Kant. (Ibid., p. 167)

For Rorty:

> [D]oubts … about Russell's notion of "knowledge by acquaintance" … came to a head … in the early 1950s, with the appearance of Wittgenstein's *Philosophical Investigations*, Austin's mockery of "the ontology of the sensible manifold," and Sellars's "Empiricism and the Philosophy of Mind". (Ibid., p. 169)

Further:

> [I]n the early fifties, Quine's "Two Dogmas of Empiricism" challenged this distinction [between "true by virtue of meaning" and "true by virtue of experience"], and with it the standard notion (common to Kant, Husserl, and Russell) that philosophy stood to empirical science as the study of structure to the study of content. Given Quine's doubts (buttressed by similar doubts in Wittgenstein's *Investigations*) …, it became difficult to explain in what sense philosophy had a separate "formal" field of inquiry, and thus how its results might have the desired apodictic character. (Ibid.)

According to Rorty, these challenges to Russellian views "were challenges to the very idea of a 'theory of knowledge,' and thus to philosophy itself, conceived of as a discipline which centers around such a theory" (ibid.).

Somewhat similarly, in his 2000 book *Articulating Reasons*, Robert Brandom writes:

> The later Wittgenstein, Quine, and Sellars (as well as Dummett and Davidson) are linguistic pragmatists, whose strategy of coming at the meaning of expressions by considering their use provides a counterbalance to the Frege-Russell-Carnap-Tarski platonistic model-theoretic approach to meaning, (Brandom 2000, pp. 6–7)

thus presenting Russell, as privileging reference over use in "the order of semantic explanation" (ibid., p. 1), an order of explanation that is reversed by Wittgenstein, Quine, and Sellars. Again, in his 2004 contribution to the *Cambridge Companion to Quine*, Hylton begins his paper "Quine on Reference and Ontology", by writing:

> Let us begin with the views of Russell, which form a sharp and useful contrast with those of Quine on these topics. Russell postulated a direct and immediate relation between the mind and entities outside the mind, a relation he called acquaintance; this relation he held to lie at the base of all knowledge. … For Russell … reference, a version of the relation between a name and the named object, is a presuppositionless relation that is at the foundation of all knowledge. Quine, as we shall see, rejects the Russellian view utterly, in every aspect, (Hylton 2004, pp. 115–6)

presenting Russell's views of acquaintance and reference as standing in stark opposition to Quine's.

Recently, there has been a growing awareness that Russell's post-1918 writings call into question the view of Russell as a philosopher whose guiding assumptions are, for better or worse, are rejected by such later figures Quine, Sellars, and the later Wittgenstein. For an examination of those writings shows that by the early 1920s Russell himself was advocating views—including an anti-foundationalist naturalized epistemology, an account of thought that rejects acquaintance as a "direct and immediate relation between the mind and entities outside the mind", and a behaviorist-inspired account of what is involved in understanding language—that are more typically associated with philosophers from later decades whom Rorty, among others, presents as dismantling Russell's philosophy. Thus, for example, Thomas Baldwin begins his 2003 paper "From Knowledge by Acquaintance to Knowledge by Causation" by writing:

> There are many familiar themes in Russell's repertoire, but his later discussions of knowledge include many insights which have received little notice. Indeed, it is often supposed that in the years after 1914, after the heroic foundational phase of analytical philosophy celebrated in countless anthologies, Russell ceased to engage in creative philosophy…. One thing I want to show here is that during these years Russell was in fact developing a new conception of epistemology, linked to a new philosophy of mind, which was so far ahead of its time that it passed by largely unappreciated. It is only now that our that our own philosophy of mind has caught up with the 'naturalisation' of the mind that Russell was teaching from 1921 onwards that we can recognise in his later writings the central themes of our current debates…. (Baldwin 2003, p. 420)

Similarly, in his 1996 paper "Quine and Wittgenstein: The Odd Couple", Burton Dreben writes:

> By late spring of 1918, Knowledge by Acquaintance together with The Knowing Subject—the very core of what had been (Analytic) Epistemology for Russell—disappear. … For the first time the nature of language *per se* is on centre stage, and Russell seeks a naturalist, indeed physicalist and broadly behaviorist account of it and of all other so-called mental activities. (Dreben 1996, p. 48)

Numerous passages support these claims of Baldwin and Dreben. Thus, for example, in his 1924 paper "Logical Atomism", Russell writes:

> I began to think it probable that philosophy had erred in adopting heroic remedies for intellectual difficulties, and that solutions were to be found merely by greater care and accuracy. This view I have come to hold more and more strongly as time went on, and it has led me to doubt whether philosophy, as a study distinct from science and possessed of a method of its own, is anything more than an unfortunate legacy from theology, (Russell 1924, p. 163)

thus anticipating the sort of "naturalism" reflected in Quine's remark that "I see philosophy not as an *a priori* propaedeutic or groundwork for science, but as continuous with science" (Quine 1969, p. 126). In "On Vagueness", Russell writes:

> My own belief is that most of the problems of epistemology, in so far as they are genuine, are really problems of physics and physiology; moreover, I believe that physiology is only a complicated branch of physics, (Russell 1923, p. 154)

thus endorsing a "naturalized epistemology" consistent with Quine's view that "epistemology in its new [naturalized] setting … is contained in natural science, as a chapter of psychology" (Quine 1969, p. 83). Again, in 1920, Russell writes

> [I am] one who regards thought as merely one among natural processes, and hopes that it may be explained some day in terms of physics. … For my part, I do not regard the problem of meaning as one requiring such special methods as are commonly called "philosophical". I believe that there is one method of acquiring knowledge, the method of science; and that all specially "philosophical" methods serve only the purpose of concealing ignorance. … Now meaning is an observable property of observable entities, and must be amenable to scientific treatment. My object has been to endeavour to construct a theory of meaning after the model of scientific theories, not on the lines of traditional philosophy, (Russell 1920, pp. 90–1)

thus anticipating Quine's view that "[t]here is nothing in linguistic meaning beyond what is to be gleaned from overt behavior in observable circumstances" (Quine 1992, p. 38). Further, in his 1921 book *The Analysis of Mind*, Russell writes:

> We may say that a person "understands" a word when (a) suitable circumstances make him use it, (b) the hearing of it causes suitable behavior in him. We may call these two active and passive understanding respectively. Dogs often have passive understanding of some words, but not active understanding, since they cannot use words.
>
> It is not necessary, in order that a man should "understand" a word, that he should "know what it means," in the sense of being able to say "this word means so-and-so." Understanding words does not consist in knowing their dictionary definitions, or in being able to specify the objects to which they are appropriate. … Understanding language is more like understanding cricket: it is a matter of habits, acquired in oneself and rightly presumed in others. To say that a word has a meaning is not to say that those who use the word correctly have ever thought out what the meaning is: the use of the word comes first, and the meaning is to be distilled out of it by observation and analysis. Moreover, the meaning of a word is not absolutely definite: there is always a greater or less degree of vagueness. (Russell 1921a, pp. 197–8)

Here, by writing that "the use of the word comes first, and the meaning is to be distilled out of it", Russell employs what Brandom calls the "strategy of coming at the meaning of expressions by considering their use" and anticipates Wittgenstein's view that "[w]e are inclined to forget that that it is the particular use of a word only which give the word its meaning" (Wittgenstein 1958, p. 69).[3] Further, by claiming

[3]Griffin (2003, p. 35, note 45) is one of the few to suggest that Russell's views of meaning and understanding in *The Analysis of Mind* had a positive influence on Wittgenstein: "It seems quite possible that *The Analysis of Mind* was the original source of Wittgenstein's view that meaning is use."

that the meaning of a word we are able to "distill" out of its use "is not absolutely definite" but rather admits of "a greater or lesser degree of vagueness", Russell appears to advocate an indeterminacy thesis of the sort that Quine articulates when he writes, for example: "When … we turn thus toward a naturalistic view of language and a behavioral view of meaning, … [w]e give up an assurance of determinacy" (Quine 1969, p. 28).

Such apparent similarities between Russell's later views and those of his supposed "heretical followers" raise the question as to why the later Russell has been written out of "the story" of analytic philosophy in favor of others whose views he anticipates by decades. My concern here, however, is to address, not this question, but rather a suggestion made by some who are aware of Russell's post-1918 views —namely, that while the later Russell does not conform to the figure of the "foundationalist", "Augustinian" Russell, the pre-1918 Russell does. Thus, for example, in the passage I have quoted above, Baldwin suggests that Russell's role in "the heroic foundational phase of analytical philosophy" is a "familiar theme" that "is celebrated in countless anthologies", and he adds later:

> Russell's (1918) imprisonment marks his transformation from the familiar author of *Principia Mathematica* to the unfamiliar author of *The Analysis of Mind* and his subsequent writings. The key change is a new determination to bring science into philosophy …. (Baldwin 2003, p. 439)

Again, in a footnote to the passage I have quoted above in which he presents "the views of Russell" as "form[ing] a sharp and useful contrast with those of Quine", Hylton writes:

> I shall speak here of Russell's views in the first two decades of this [twentieth] century and shall not be concerned with subsequent shifts of doctrine. (Hylton 2004, p. 146, note 1)

Thus he suggests that while Russell's views by 1920 may not provide "a sharp and useful contrast with those of Quine", Russell's views before then do. And in a footnote to the passage I have quoted above in which he indicates that by the spring of 1918, "the nature of language *per se* is on centre stage, and Russell seeks a naturalist, indeed physicalist and broadly behaviorist account of it and of all other so-called mental activities", Dreben writes:

> For the pre-linguistic Russell, see P. Hylton, *Russell, Idealism and the Emergence of Analytic Philosophy*, (Dreben 1996, p. 58, note 33)

thus suggesting that he endorses Hylton's understanding of Russell's views prior to his 1918 linguistic turn.

In what follows I argue that "the pre-linguistic Russell" is not a monolithic figure conforming to the stereotype of the "Augustinian" Russell: while the Moorean Russell does, the post-Peano Russell—in particular, the Russell, who in *The Principles of Mathematics* presents analyses of, for example, the cardinal numbers and the real numbers, and who defends Cantor's analyses of infinity and continuity—does not, so that shortly after the Paris Congress of 1900, Russell himself begins to undermine views with which he is often most closely associated. In particular, I focus

on differences between Russell's view of analysis before and after the Paris Congress and the role that vagueness comes to play in his conception of analysis. I begin by introducing views that are central to the Moorean conception of analysis, and discuss, in particular, how these views are reflected in Russell's Moorean discussion of theories of order, including the theory of number. After showing that, in the course of completing *The Principles of Mathematics*, Russell comes to accept definitions of cardinal number that undermine his Moorean theory of number, and, more generally Moorean theories of order, I argue that in defending these definitions of the cardinal numbers, as well as in other analyses of mathematical concepts he presents in *The Principles of Mathematics*, Russell does not apply the Moorean conception of analysis, but rather introduces the notion of vagueness—the same notion of vagueness that he presents in his 1923 paper—and characterizes analysis as a matter of making "precise" or "definite" what was previously vague. After discussing some differences between Russell's views before and after his 1918 prison stay, I conclude by arguing against the way in which Hylton characterizes Russell's pre-prison view of analysis and contrast it with Quine's.

The Moorean Russell and Analysis

I argue now that the Moorean Russell accepts both an "Augustinian" view of meaning, according to which

(Aug) The meaning of a word is an entity (simple or complex) corresponding to that word; the meaning of a sentence is a proposition—a complex entity—whose constituents are the meanings of the words in that sentence

as well as the "principle of acquaintance", according to which

(PoA) Apprehending a proposition requires being acquainted with each of its constituents,

and that accepting this combination of views places a strict condition on analysis.

In attributing these views to the Moorean Russell, I am not thereby denying that there are passages in which Russell endorses them in his post-Peano writings. Indeed, as I indicate, some of the most well-known passages in which Russell endorses (Aug) and (PoA) occur in post-Peano writings. Rather, my claim in this section is merely that Russell comes to accept these views during his early post-Idealist, pre-Peano Moorean period, not that he ceases to endorse them immediately after the Paris Congress. Below, I argue that, despite these post-Peano passages in which he continues to endorse (Aug) as well as (PoA), Russell's post-Peano practice of analysis of mathematical concepts and his post-Peano characterization of that practice of analysis are not in accord with the conception of analysis that follows from accepting (Aug) along with (PoA). More specifically, I argue that in coming to emphasize the role that vagueness plays in his conception of analysis, Russell regards ordinary, as opposed to "precise", language as failing to

comply with (Aug) and thereby has a way to accept (PoA), without accepting the Moorean conception of analysis. In contrast, in his post-1918 writings, Russell not only denies that any language is "precise" (thereby denying that any language is meaningful by the standard of (Aug)) but has also rejected the notion of acquaintance (and hence (PoA)).

First of all, in accord with (Aug), Russell writes in 1899:

> Philosophically, a term is defined when we are told its *meaning*…. It will be admitted that a term cannot be usefully employed unless it means something. What it means is either complex or simple. That is to say, the meaning is either a compound of other meanings, or it is itself one of those ultimate constituents out of which other meanings are built up. In the former case, the term is philosophically defined by enumerating its simple constituents. But when it is itself simple, no philosophical definition is possible. [Here, as elsewhere, emphasis is in the original unless noted otherwise.] (Russell 1899, p. 410)

Thus, he makes the "Augustinian" move of assimilating a word's having meaning —being "usefully employed"—to its standing for an entity, simple or complex. Moore makes the same move in *Principia Ethica*, where he writes:

> [I]f it is not the case that 'good' denotes something simple and indefinable, only two alternatives are possible: either it is a complex … or else it means nothing at all, and there is no such subject as Ethics, (Moore 1903a, p. 15)

thereby indicating that for a word (here "good") is to be meaningful at all, it must "denote" an entity—"something"—simple or complex.

Likewise, Russell reflects a commitment to (Aug) in *The Principles of Mathematics*, where after writing:

> [I]t must be admitted, I think, that every word occurring in a sentence must have *some* meaning: a perfectly meaningless sound could not be employed in the more or less fixed way in which language employs words, (Russell 1903, p. 42)

he adds five pages later:

> *Words* all have meaning, in the simple sense that they are symbols which stand for something other than themselves. But a proposition, unless it happens to be linguistic, does not itself contain words: it contains the entities indicated by words. (Ibid., p. 47)

Thus, as in 1899, Russell makes a seamless transition from indicating that "every word occurring in a sentence" must be meaningful—that is, "must have *some* meaning"—to indicating that its thus having a meaning consists in its standing for an entity.

Further, although the Moorean Russell does not use the word "acquaintance", the notion for which he later uses that term is central to his philosophy. For Moore and the Moorean Russell, to be "conscious of" an object—either by sensing it or by thinking of it—is to "know" or to be "directly aware" of that object; it is to have the

object "present to one's mind".[4] In Russell's later terminology, it is to be acquainted with an object. Hence, when the Moorean Russell writes in 1900:

> [E]verything that can occur in a proposition must be something more than a *mere* idea—it must be the object of an idea, *i.e.* an entity to which an idea is related …. On the other hand, whatever can form part of a judgment which we make must be the *object* of one of our ideas, (Russell 1900b, p. 229)

he is endorsing (PoA), the view he expresses in 1905, in the penultimate paragraph of "On Denoting", where he writes:

> [I]n every proposition that we can apprehend …, all the constituents are really entities with which we have immediate acquaintance, (Russell 1905, p. 427)

the view he similarly expresses in 1911 in "Knowledge by Acquaintance and Knowledge by Description" as well as in 1912, in *The Problems of Philosophy*, when he writes:

> Every proposition which we can understand must be composed wholly of constituents with which we are acquainted. (Russell 1911a, p. 154, 1912, p. 58)[5]

For the Moorean Russell, that is, and for the Russell who defends (PoA), propositions do not, in general, have as their constituents mental entities; however, apprehending a proposition requires that constituents of that proposition be "objects" of our "ideas"—that is, that they be objects with which we are acquainted, objects "present to the mind". That we can thus apprehend propositions requires—what is central to both Moore and Russell—that the mind can have direct cognitive contact with entities that are not themselves mental entities.

Strictly speaking, (PoA) is not a view regarding language: it is a view regarding our apprehension of certain non-linguistic entities—namely, propositions. However, the combination of (Aug) and (PoA) naturally leads to a view regarding what is required to understand a sentence. For by (Aug), the meaning of a sentence is the proposition it expresses, a complex entity whose constituents are the meanings of the words in that sentence. Hence, insofar as understanding a sentence requires knowing its meaning, by (Aug), understanding a sentence requires knowing, or apprehending, the proposition it expresses, in which case, by (PoA), understanding a sentence requires being acquainted with each constituent of that

[4]This is the central claim of Moore (1903b), the main argument of which Russell (1912, Chap. 4) repeats, using the word "acquaintance".

[5]By 1910, in accepting his multiple-relation theory of judgment (MRTJ), Russell no longer holds that there are entities that are propositions. The sense in which he then accepts (PoA) is that he holds that "whenever a relation of supposing or judging occurs, the terms to which the supposing or judging mind is related by the relation of supposing or judging must be terms with which the mind in question is acquainted" (Russell 1911a, p. 155). Here, as I discuss in more detail below, "the terms to which the supposing or judging mind is related" are (in general) non-linguistic entities, entities that as Russell writes in his 1910 paper "The Theory of Logical Types" (where he first endorses the MRTJ) "are called the constituents of the proposition" (Russell 1910, pp. 10–1).

proposition—that is, requires knowing, or being acquainted with (what, by (Aug), is) the meaning of each word in a sentence expressing that proposition. And in *The Problems of Philosophy*, Russell presents (PoA) as an obvious consequence of the Augustinian view of meaning:

> We must attach *some* meaning to the words we use, if we are to speak significantly and not utter mere noise; and the meaning we attach to our words must be something with which we are acquainted. (Russell 1912, p. 58)

Given, that is, that the meaning of a word is an entity corresponding to that word, then, for Russell, understanding a sentence containing that word requires knowing—in particular, being acquainted with—the entity that is the meaning of that word. And this is, in effect, what accepting (PoA) in the context of accepting (Aug) commits one to—namely, that understanding a sentence requires knowing—being acquainted with—the entities that are the meanings of words in that sentence. Thus, in his 1913 *Theory of Knowledge* manuscript, Russell is applying (PoA) in the context of a sentence whose words are meaningful by the standard of (Aug) when he writes:

> Let us take as an illustration some very simple proposition, say "*A* precedes *B*", where *A* and *B* are particulars. In order to understand this proposition, it is … obviously necessary that we should know what is meant by the words which occur in it, that is to say, we must have acquaintance with *A* and *B* and with the relation "preceding". (Russell 1913, pp. 110–1)

Accepting (Aug) along with (PoA) places a severe constraint on what is involved in analyzing the proposition expressed by a given sentence. For by (Aug) together with (PoA), given a sentence *S*, its meaning is a proposition *P*, and understanding *S* requires being acquainted with each constituent of *P*. In that case, analyzing *P*—a task that involves identifying its ultimate constituents—involves making explicit the entities with which anyone who understands *S* is acquainted. That is, it will involve making explicit what is already "present to the mind" of one who, prior to analysis, understands *S* (or, indeed, any sentence expressing *P*).

Accordingly, applying a conception of analysis that incorporates (Aug) along with (PoA) may lead one to reject analyses on the basis that they fail to accord what is "before our minds" when we understand the expressions in question.[6] In *Principia Ethica*, Moore, in effect, applies this conception of analysis when he writes:

> Every one does in fact understand the question 'Is this good?' When he thinks of it, his state of mind is different from what it would be, were he asked 'Is this pleasant, or desired, or

[6]Thus the Moorean conception of analysis invites the so-called "paradox of analysis" as to how, or in what sense, the analysis of a proposition expressed by a sentence we understand can be informative, if it can reveal nothing with which we were not already acquainted in our original understanding of that sentence (see Moore 1942, 665f). What I point out here, in effect, is that in applying the Moorean conception of analysis, the early Moore and Russell are often not attempting to provide apparently "informative" analyses of familiar concepts, but rather argue against apparently "informative" analyses provided by others, on the basis that such analyses conflict with what is "present to the mind" of one who understands the words in question.

> approved?' It is has distinct meaning for him …. Whenever he thinks of 'intrinsic value,' or 'intrinsic worth,' or says that a thing 'ought to exist,' he has before his mind the unique object —the unique property of things—which I mean by 'good'. Everybody is constantly aware of this notion…. (Moore 1903a, pp. 16–7)

Thus he indicates that because we have different entities "before the mind" when we understand the questions "Is this good?" and "Is this pleasant?", then these sentences express different propositions, in which case it is wrong to define what is good as what is pleasant. Likewise, in 1909, Russell applies this conception of analysis in arguing against pragmatist accounts of "the meaning of truth". In particular, after writing generally:

> When we ask "what does such and such a word mean?" what we want to know is "what is in the mind of a person using the word?"

he adds in the following paragraph:

> When we say that a belief is true the thought we wish to convey is not the same thought as when we say that the belief furthers our purposes; thus "true" does not mean "furthering our purposes"…. Thus pragmatism does not answer the question: What is in our minds when we judge that a certain belief is true? (Russell 1909, p. 274)

For Russell, since what is "present" to our minds when we understand "That belief is true" and "That belief furthers our purposes" is not the same, then, given (Aug) along with (PoA), these sentences express different propositions and *being true* should not be analyzed as *furthering our purposes*. Similarly, in his 1910 paper "The Theory of Logical Types", Russell writes:

> [A] convenient way to read "(x). ϕx" is "ϕx is true for all possible values of x". This is, however, a less accurate reading than "ϕx always", because the notion of *truth* is not part of the content of what is judged. When we judge "all men are mortal", we judge truly, but the notion of truth is not necessarily in our minds, any more that it need be when we judge "Socrates is mortal". (Russell 1910, p. 8)

Here, Russell indicates, in effect, that because what is present to our minds when we understand, for example, "Socrates is mortal" and "That Socrates is mortal is true", is different, these sentences do not express the same "content", nor for similar reasons do corresponding sentences of the forms "(x). ϕx" and "ϕx is true with all possible values of x".

During his Moorean period, one of Russell's main concerns is the nature of order. He distinguishes generally between absolute and relative theories of order and applies that distinction to theories of time, magnitude, and number, as well as space (which, is more complicated since it involves more than one dimension), colors, and pitches of sounds. As I discuss now, the way in which the Moorean Russell decides between these competing theories of order reflects his commitment to the Moorean conception of analysis.

Theories of Order and Moorean Analysis

For Russell absolute and relative theories of order present competing metaphysical accounts—different views of the "indefinables" or "simples" or "ultimate constituents of the universe"— for the sort of order in question. Thus, for example, early in his paper "Is Position in Time Absolute or Relative?", which he delivered in May 1900, Russell contrasts the relative and absolute theories of time by writing:

> Does an event occur at a time, or does it merely occur before certain events, simultaneously with others, and after a third set? The relational theory of time holds the latter view …. The absolute theory, on the contrary, holds that events occur *at* times, that times are before or after each other, and that events are simultaneous or successive according as they occur at the same or different times. (Russell 1900b, p. 222)

Thus, on the relational theory, the only indefinable terms to be related are events, so that "times do not really exist" (Russell 1901b, p. 242), and there are three indefinable temporal relations that may obtain between events—*before*, *after*, and the symmetric, transitive relation *simultaneity*. On the absolute theory, in contrast, there are both events and absolute moments among the ultimate constituents of the universe. Here, moments have an "intrinsic order" to one another, while events acquire a temporal order only "by correlation" with the "independent" or "self-sufficient" series of moments in absolute time (see Russell 1901a, p. 291). Here too there are three primitive relations—*before* and *after*, now understood as relations between moments not events, and *occurring at* which relates an event to the moment at which it occurs.

For Russell, one way to focus the difference between the two theories is to consider the analysis of the proposition expressed by an instance of

$$(\text{Time}_1) \quad \text{Event } \alpha \text{ is simultaneous with event } \beta.$$

In accord with (Aug), Russell holds that that the expression "is simultaneous with" has as its meaning an entity—a relation, namely *simultaneity*. However, for Russell, the central philosophical issue is whether that relation is an indefinable, an ultimate constituent of the universe, or whether it is definable. On the relative theory of time, that relation is indefinable, so that an instance of (Time_1), expresses a proposition that has three ultimate constituents—namely, the events in question and *simultaneity*. In contrast, on the absolute theory, *simultaneity* is to be analyzed in terms of *occurring at* the same moment, so that the full analysis of a proposition expressed by an instance of (Time_1) is given by the corresponding instance of

$$(\text{Time}_2) \quad \text{there is a moment } t \text{ such that } \alpha \text{ occurs at } t \text{ and } \beta \text{ occurs at } t.$$

As Russell writes, on the absolute theory

> "*A* is simultaneous with *B*" requires analysis into "*A* and *B* are both at one time". (Russell 1899–1900, p. 147)

Likewise, for Russell, the central issue distinguishing the relative theory of magnitude from the absolute theory is whether, when two quantities are equal in magnitude, there is some further indefinable entity—a magnitude—that is common to the two quantities. On the relative theory of magnitude, there is no such indefinable magnitude and the (transitive, symmetrical) relation of *equality in magnitude* is an indefinable relation between quantities, so that an instance of

(Mag_1) Quantity α is equal in magnitude to quantity β

expresses a proposition that has three ultimate constituents—the quantities α and β and the indefinable relation of *equality in magnitude*. In contrast, on the absolute theory of magnitude, the relation of *equality of magnitude* is definable, so that

(Mag_2) There is a magnitude m such that α has m and β has m

is a perspicuous representation of the proposition expressed by the corresponding instance of (Mag_1). As Russell writes in his pre-Peano 1899–1900 draft of PoM:

> The kernel of the difference between the present [relative] theory and the former [absolute theory] is, that now equality is taken as indefinable, whereas formerly each magnitude was indefinable. … It might perhaps be thought, by those who regard definition as subject to convenience, that the present theory is not incompatible with the former. This would, however, be a grave philosophical error. Every concept is necessarily either simple or complex, and it is not in our power to alter its nature in this respect. If it is complex, it should be analyzed and defined; if simple, it should be used in defining other terms, without itself receiving a definition. Thus equality either may be analyzed into sameness of magnitude, or it may not be so analyzed. … It does not lie with us to choose what terms are to be indefinable; on the contrary, it is the business of philosophy to discover these terms. We have to decide whether the indefinable term is the relation of equality, or a common property of equal quantities. If we choose the former alternative, we shall have to deny a common property; for if there were any common property, this could be used to define equality. (Ibid., pp. 57–8)

For the Moorean Russell, in deciding which of these theories is correct, we are not concerned with matters of convenience or with arriving at a theory which has the fewest indefinables; rather, we are attempting to determine what are among the indefinable, ultimate constituents of the universe. Are there, in addition to quantities, indefinable magnitudes? Or, are there no such magnitudes, but instead an indefinable symmetric transitive relation of *equality in magnitude*?

Again, in distinguishing the relative from absolute theory of number, Russell indicates that whereas

> [T]he relational theory would hold that there is never a number of terms at all, but there are merely the relations of equal, greater, and less among collections, [whereas on the absolute theory] equality … consists in possession of the same number. (Russell 1900b, p. 225)

Thus, on the relative theory of number, there are no indefinable numbers in addition to "collections", so that the proposition expressed by an instance of

$$(\text{Num}_1)\ \text{Class}\ \alpha\ \text{is equal in number with class}\ \beta$$

has as among its constituents an indefinable relation of *being equal in number*. In contrast, on the absolute theory of number, there are indefinable numbers but no indefinable relation of *being equal in number*, so that an instance of

$$(\text{Num}_2)\ \text{There is a number}\ n\ \text{such that}\ \alpha\ \text{possesses}\ n\ \text{and}\ \beta\ \text{possesses}\ n.$$

is the privileged representation of the proposition expressed by the corresponding instance of (Num_1).

More generally, for Russell, deciding between relative and absolute theories of order requires considering instances of

$$(\text{Ab}_1) \quad E(\alpha, \beta),$$

where E is a symmetrical transitive relation (such as *simultaneity*, *equality in magnitude*, or *equality in number*), and

$$(\text{Ab}_2) \quad (\exists x)(R(\alpha, x)\ \&\ R(\beta, x)),$$

where R is an appropriate many-one relation (such as *occurring at*, *having*, or *possessing*). To accept a relative theory of order is to hold that the relevant transitive, symmetrical relation is indefinable so that the relevant instances of (Ab_1) express propositions that contain that relation as an ultimate constituent. In contrast, to accept the corresponding absolute theory of order is to hold that that symmetrical transitive relation is definable, and that the relevant instances of (Ab_2) are privileged representations of the propositions expressed non-perspicuously by corresponding instances of (Ab_1).

During his Moorean period and even in his draft of *The Principles of Mathematics* immediately following the Paris Congress, Russell accepts absolute theories of order. Moreover, in accord with accepting (PoA) along with (Aug), Russell indicates that we are in a position to settle the issue by immediate "inspection". Thus, in his pre-Peano draft of *The Principles of Mathematics*, Russell defends the absolute theory of magnitude by writing: "[W]hen we consider what we mean when we say that two quantities are equal, it seems preposterous to maintain that they have no common property not shared by unequal quantities" (Russell 1899–1900, p. 58). In accord with accepting (Aug) along with (PoA), that is, Russell claims that if we simply consider "what we mean" when we utter a sentence of the form (Mag_1), we will recognize that the proposition expressed does not have an indefinable relation of *equality in magnitude* as an ultimate constituent and that that proposition is instead perspicuously represented by the corresponding instance of (Mag_2).

Likewise, in defending the absolute theory of time, he writes: "A direct consideration of the question … makes it very difficult to hold that simultaneous events have absolutely nothing in common beyond the common qualities of all events"

(Ibid., p. 227); and in defending the absolute theory of number, he writes that numerical "equality [in number] plainly consists in possession of the same number" (ibid., p. 225; see also Russell 1899–1900, p. 146). Further, he writes: "In cases where, as with numbers and colours, these positions [that is, the absolute positions in independent series] have names, the absolute theory is plainly correct." (Russell 1900b, p. 226) Here, Russell suggests, in accord with accepting (Aug) along with (PoA), that in understanding the name of a number or color, we will thereby know—be acquainted with—the indefinable, ultimate constituent of the universe it stands for.

More generally, he writes: "[F]or my part I consider it self-evident that all symmetrical transitive relations are analyzable" (Russell 1901c, p. 262); and in his post-Peano draft of *The Principles of Mathematics* he incorporates what he thus regards as "self-evident" into an "axiom" according to which the full analysis of an instance of (Ab_1) is given by the corresponding instance of (Ab_2):

> [M]y axiom of abstraction, which precisely stated, is as follows: "Every transitive symmetrical relation, of which there is at least one instance, is analyzable into joint possession of a new relation to a new term, the new relation being such that no term can have this relation to more than one term, but that its converse does not have this property." (see Russell 1903, p. 220, as correlated with Byrd 1996–7, pp. 165–6)

In all these passages, Russell is indicating that the metaphysics of time and number, and more generally of order, can be settled by appeal to what is obvious by "direct consideration" of what we mean by sentences of the form (Ab_1). In merely reflecting what is involved in understanding such a sentence, we will recognize that the proposition expressed by such a sentence is given, not by a sentence of the form (Ab_1), but rather by the corresponding sentence of the form (Ab_2).

Russell's views here, incidentally, are in accord with those of both Moore and Husserl. Thus, for example, in his 1901 paper "Identity", Moore writes that on his view it is correct "to *define* the relation of exact similarity between two things as involving relation to a third thing" (Moore 1901, p. 131). Likewise, in his *Logical Investigations*, Husserl writes:

> [W]e find … that wherever things are 'alike', an identity in the strict and true sense is also present. We cannot predicate exact likeness of things, without stating the respect in which they are thus alike. Each exact likeness relates to a Species, under which the objects compared are subsumed …. It would of course appear as a total inversion of the true state of things, were one to try to define identity, even in the sensory realm, as being essentially a limiting case of 'alikeness'. Identity is wholly indefinable, whereas 'alikeness' is definable: 'alikeness' is the relation of objects falling under one and the same Species. If one is not allowed to speak of the identity of the Species, of the respect in which there is 'alikeness', talk of 'alikeness' loses its whole basis. (Husserl 1900–1, p. 242)

Thus, in accord with Russell's "axiom of abstraction", both Moore and Husserl hold that whenever a proposition is expressed by a sentence of the form "α is exactly similar (in a given respect) to β", that proposition is more perspicuously represented by a sentence indicating that the objects in question bear the same relation (for Husserl, the relation of "falling under") to the a third entity (for

Husserl, a "Species"). Indeed, as late as 1911, in lectures that were later published as *Some Main Problems of Philosophy*, Moore writes:

> Consider … the group formed of all … pairs or couples. … The property [which all and only couples share] does seem to consist in the fact that *the number two* belongs to every such collection and only to such a collection …. And it seems to me that … we can hold the number two before our minds, and see what it is, and *that* it is, in almost the same way as we can do this with any particular sense-datum we are directly perceiving, (Moore 1953, p. 366)

thereby indicating not only that he accepts, in Russell's terms, the absolute theory of numbers, but also that he is acquainted with the entities that, on that theory, are the numbers.

From the "Axiom of Abstraction" to the "Principle that Dispenses with Abstraction"

Following the Paris Congress of August 1900, Russell drafts his paper "The Logic of Relations" in October and writes final drafts of Parts III–VI of *The Principles of Mathematics* in November and December 1900. During this period, he holds—what he had formerly denied—that Dedekind, Weierstrass, and, especially Cantor, had solved all the traditional problems of infinity and continuity, but he does not yet accept the so-called "Frege-Russell" definitions of cardinal numbers. However, in the spring of 1901, in producing the final version of "The Logic of Relations", Russell accepts the following definition of the cardinal number of class α:

$$(\mathrm{Num}_{\mathrm{df}}) \quad \text{The cardinal number of } \alpha =_{\mathrm{df}} \{\xi : \xi \text{ is similar to } \alpha\},$$

where two classes are similar, or equal in number, if and only if the members of those classes can be put into a one-to-one correspondence with each other. As Russell writes in *The Principles of Mathematics*, "we decide to identify the number of a class with the whole class of classes similar to the given class" (Russell 1903, p. 305; see also p. 115). However, even after introducing ($\mathrm{Num}_{\mathrm{df}}$), and up until copyediting the page proofs of *The Principles of Mathematics*, Russell holds, in accord with his Moorean absolute theory of numbers, that "philosophically", if not "formally", numbers are indefinable. Thus, in his final draft of Part II of *The Principles of Mathematics* composed, apparently, in May 1902, Russell writes that "for formal purposes, numbers may be taken to be classes of similar classes", but he then produces an argument intended to show that

> Numbers, it would seem, are … philosophically, not formally indefinable. … [T]hese indefinable entities are different from the classes of classes which it is convenient to call numbers in mathematics. (Byrd 1987, p. 69)

Hence, it is only sometime after June 1902, in his final copyediting of *The Principles of Mathematics*, that Russell changes this passage to read:

> Numbers are classes of classes, namely of all classes similar to a given class. … [N]o philosophical argument could overthrow the mathematical theory of cardinal numbers set forth [above]. (Russell 1903, p. 136)

Hence, as late as May 1902, Russell distinguishes each cardinal number from the class of similar classes he identifies it with in *The Principles of Mathematics* as published.

In coming to hold that it is philosophically acceptable to regard cardinal numbers as classes of similar classes, Russell has rejected the "absolute" theory of number. No longer are numbers ultimate constituents of the universe constituting a domain separate from classes; now they simply are classes of similar classes. No longer are numbers regarded as indefinable entities in terms of which the relation of *similarity* (or *being equal in number*) between classes is defined; now, that relation of *similarity* is used to define number.[7] No longer is the proposition expressed by an instance of (Num_1) perspicuously represented by the corresponding instance of (Num_2). For, given (Num_{df}), to say of two classes that they possess the same cardinal number is really to say that the class of classes similar to the first is identical to the class of classes similar to the second, so that an instance of (Num_2) is not a perspicuous representation of the proposition it expresses, but rather expresses a proposition that is more[8] perspicuously represented by the corresponding instance of

$$(\text{Num}_3) \quad \{\omega : \omega \text{ is similar to } \alpha\} = \{\omega : \omega \text{ is similar to } \beta\}.$$

Further, since on Russell's view in *The Principles of Mathematics*, to say of two classes that they are similar is not to invoke the notion of cardinal number, either as indefinables or as classes of similar classes but is rather to say that there is a one-to-one correspondence between the members of those classes, an instance of (Num_1) expresses a distinct proposition from that expressed by corresponding instances of (Num_2) and (Num_3). Thus, Russell has gone from holding that corresponding instances of (Num_1) and (Num_2) express the same proposition—where the latter, but not the former, do so perspicuously—while corresponding instances of (Num_3) express distinct propositions (since indefinable numbers are different from classes of similar classes) to holding that corresponding instances of (Num_2) and (Num_3) express the same proposition—where the latter, but not the former, do so perspicuously—while corresponding instances of (Num_1) express distinct propositions.

Moreover, in *The Principles of Mathematics*, Russell recognizes that he can introduce definitions similar to (Num_{df}) that will likewise undermine other

[7] Although here, unlike the relative theory of number he considered during his Moorean period, similarity is not indefinable but is rather defined in terms of one-to-one correspondence.

[8] "More" perspicuously, because a fully perspicuous representation of that proposition would have to take into account the definition of "similarity" in terms of one-to-one correspondence.

"absolute" theories of order. In particular, in the course of defending his definitions of cardinal numbers, Russell writes:

> Wherever Mathematics derives a common property from a reflexive, symmetrical, and transitive relation, all mathematical purposes of the supposed common property are completely served when it is replaced by the class of terms having the given relation to a given term; and this is precisely the case presented by cardinal numbers. (Ibid., p. 116)

Here, Russell indicates that whenever one holds—as his former "axiom of abstraction" requires—that a symmetric transitive relation indicates the possession of a "common property", one may "replace" the "supposed common property" by "the class of terms have the given [symmetric transitive] relation to a given term". As Russell indicates, this is exactly the procedure he follows in introducing (Num_{df}): instead of regarding the cardinal number of a given class α as an indefinable "property" common to α and any class similar to α, regard it as the class of "terms" (here, classes) having the given relation (here, *similarity*) to the given term (here, class α). In the case of magnitudes, one would introduce

$$(\text{Mag}_{\text{df}}) \quad \text{The magnitude of quantity } q = \{x : x \text{ is equal in magnitude to } q\},$$

so that the magnitude of a given quantity q is regarded, not as an indefinable "property" common to q and any quantity equal in magnitude to q, but rather as the class of quantities equal in magnitude to q. And Russell introduces this view of magnitude in a footnote he adds to Part III of *The Principles of Mathematics* in the final changes he makes proofreading the typescript (see Russell 1903, p. 167, as collated with Byrd 1996–7, p. 161). With regard to time, one would introduce

$$(\text{Time}_{\text{df}}) \quad \text{The moment at which event } e \text{ occurs } = \{x : x \text{ is simultaneous with } e\},$$

so that the moment at which event e occurs is regarded, not as an indefinable at which e and any event simultaneous with e occurs, but rather as the class of events simultaneous with e, a view of the sort he comes to develop by 1914 in which he delivers lectures published in 1915 as *Our Knowledge of the External World*.[9]

More generally, Russell is indicating in this passage from *The Principles of Mathematics* that, instead of holding—as his earlier "axiom of abstraction" requires—that an instance of (Ab_1) expresses a proposition that is perspicuously represented by the corresponding instance of (Ab_2), we can introduce, in place of the "supposed common property" quantified over in the instance of (Ab_2), the following definition:

[9]The technical difficulty in defining moments and points is that since the "events" in terms of which they are to be defined "have a finite extent", events can be "overlapping" without being entirely simultaneous, so that moments (and points) will have to be "constructed" out of "overlapping", rather than "simultaneous" events. Russell credits Whitehead with the solution of this problem (see, for example, Russell 1915, pp. 114ff, 1924, p. 166).

$$(\mathrm{Ab}_{\mathrm{df}}) \quad f(\alpha) =_{\mathrm{df}} \{x : E(x, \alpha)\},$$

where the relevant function f is related to the relevant many-one relation R in the corresponding instance of (Ab_2) so that $f(\alpha) = x$ if and only if $R(\alpha, x)$. Given such a definition, an instance of (Ab_2) is no longer a perspicuous representation of the proposition it expresses; instead, that proposition is more perspicuously represented by the corresponding instance of

$$(\mathrm{Ab}_3) \quad \{x : E(x, \alpha)\} = \{x : E(x, \beta)\},$$

while the corresponding instance of (Ab_1) expresses a distinct proposition.

Definitions of the form $(\mathrm{Ab}_{\mathrm{df}})$ eliminate certain indefinables—namely, indefinable "absolute positions" of the sort that he had previously taken numbers, magnitudes, and moments to be. And it is this point that Russell emphasizes when he discusses what he now calls "the principle of abstraction". As he writes in *Our Knowledge of the External World*:

> This principle, which might equally well be called "the principle which dispenses with abstraction," … is one which clears away incredible accumulations of metaphysical lumber. … When a group of objects have that kind of similarity which we are inclined to attribute to possession of a common quality, the principle in question shows that membership of the group will serve all the purposes of the supposed common quality, and that therefore, unless some common quality is actually known, the group or class of similar objects may be used to replace the common quality …. (Russell 1915, p. 42)

Here, in language similar to that from the passage from *The Principles of Mathematics* I have quoted above, Russell indicates that because adopting definitions of the form $(\mathrm{Ab}_{\mathrm{df}})$ enables one to avoid assuming that whenever entities bear to each other a symmetric transitive relation, there is a further indefinable entity—a "common quality"—that is common to the original entities in question, adopting such definitions "clears away incredible accumulations of metaphysical lumber".[10] What he does not mention is that he himself was one who admitted the sort of "metaphysical lumber" that he is now "dispensing with", and that in accepting definitions of the form $(\mathrm{Ab}_{\mathrm{df}})$, he is, in effect, rejecting the absolute theories of order that had been central to his Moorean philosophy.

[10]By *Principia Mathematica*, Russell dispenses with classes by accepting "no classes" theory, according to which sentences containing class symbols are interpreted so that there is no reference to any entities that are classes; however, in the passage I have quoted from *Our Knowledge of the External World*, Russell makes no allusion to his dispensing with classes and instead reflects the understanding of the "principle of abstraction" he had at the time of *The Principles of Mathematics*. Thus, as early as December 1903, Russell writes to Couturat that once he proves his earlier "axiom of abstraction" by substituting an equivalence class of objects for the "hypothical quality common to all these objects" ("*substituer la classe même des objects dont il est question à la qualité hypothétique commune à tous ces objects*"), it would be better to call the "principle of abstraction" the "principle replacing abstraction" ("*princípe remplaçant l'abstraction*"). See Schmid (2001, p. 346).

(Num_{df}), Analysis and Vagueness

In accepting (Num_{df}), Russell is not simply rejecting his early absolute theory of number; he is also rejecting the Moorean conception of analysis on the basis of which he had accepted the absolute theory of number, and, more generally, all absolute theories of order. As I have discussed, for the Moorean Russell, it is obvious that "what we mean" when we understand a sentence of the form (Num_1) is a proposition whose perspicuous representation is given by the corresponding sentence of the form (Num_2). However, when he rejects the absolute theory of number and accepts (Num_{df}), he does not likewise present his new account as being in accord with "what we mean" when we make ordinary numerical claims. That is, it is not that he continues to hold that analyzing propositions expressed by sentences containing numerical expressions requires articulating what is "present to our minds" when we understand such sentences, but changes his view as to what is, in fact, then "present to our minds". Rather, he has changed his view as to what is required of a philosophically adequate analysis, so that determining what is "present to our minds" when we understand such sentences is no longer relevant to analysis. Further, and what is related, he now presents ordinary numerical expressions as being "vague" in the sense he would articulate twenty years later, a sense according to which "vague" expression do not meet the conditions for being meaningful that are set by (Aug).

Thus, in *The Principles of Mathematics*, after introducing (Num_{df}), Russell writes:

> To regard a number as a class of classes must appear, at first sight, a wholly indefensible paradox. Thus Peano (*F*[*ormulaire de Mathématiques*], 1901, §32) remarks that "we cannot identify the number of [a class] *a* with the class of classes in question [*i.e.* the class of classes similar to *a*], for these objects have different properties." He does not tell us what these properties are, and for my part I am unable to discover them. Probably it appeared to him immediately evident that a number is not a class of classes. (Russell 1903, p. 115)

Here—and using his early terminology according to which what is "paradoxical" is not necessarily contradictory but is rather, more generally, counter-intuitive[11]—Russell acknowledges that his new account of numbers is far from obvious by direct "inspection"; on the contrary, he concedes that someone might find it "immediately evident that a number is not a class of classes".

Further, in the following paragraph he writes:

> Mathematically, a number is nothing but a class of similar classes: this definition allows the deduction of all the usual properties of numbers…. But philosophically we may admit that every collection of similar classes has some common predicate applicable to no entities except the classes in question, and if we can find, by inspection, that there is a certain class of such common predicates, of which one and only one applies to each collection of similar classes, then we may, if we see fit, call this particular class of predicates the class of

[11]Thus, for example, Russell (1903, Chap. 10) calls "the contradiction" what others typically call "Russell's paradox".

> numbers. For my part, I do not know whether there is such a class of predicates, and I do know that, if there be such a class, it is wholly irrelevant to Mathematics.… For the future, therefore, I shall adhere to the above definition, since it is at once precise and adequate to all mathematical uses. (Ibid., p. 116)

Here, Russell is, in effect, comparing his new view of cardinal numbers as classes of similar classes with his earlier absolute theory of number, according to which the cardinals are indefinable "predicates"[12] common to similar classes. He does not claim that his new view reflects more accurately than his old view "what we mean" when we make claims involving cardinal numbers. He does not even deny that there are indefinable "predicates" of classes of the sort that, on his earlier absolute theory of numbers, *are* the cardinal numbers. Instead, he claims that whether or not there are such indefinable "predicates", regarding cardinal numbers as classes of similar classes "allows the deduction of all the usual properties of numbers", in which case there is no need to address the issue as to whether there are indefinables of the sort that he previously took numbers to be.

Moreover, by writing that his definition of cardinal numbers "is at once precise and adequate to all mathematical uses", Russell is suggesting that our ordinary use of numerical expressions is not likewise "precise". Indeed, in the "Preface" to *The Principles of Mathematics*, Russell appeals to a distinction between "vagueness" and "precision" to justify many of his definitions in that work. In particular, in defending the apparent "departures from common usage" in his definitions of mathematical terms, he writes:

> As regards mathematical terms, the necessity for establishing the existence-theorem in each case—*i.e.* the proof that there are entities of the kind in question—has led to many definitions which appear widely different from the notions usually attached to the terms in question. Instances of this are the definitions of cardinal, ordinal and complex numbers. In the two former of these, and in many other cases, the definition as a class, derived from the principle of abstraction, is mainly recommended by the fact that it leaves no doubt as to the existence-theorem. But in many instances of such apparent departure from usage, it may be doubted whether more has been done than to give precision to a notion which had hitherto been more or less vague. (Ibid., p. xix)

By indicating that our ordinary notion of cardinal number is "more or less vague", Russell is indicating that, as we ordinarily use it, a numerical expression does not yet succeed in standing for any one entity that is correctly regarded as "the meaning" of that expression; and in that case, we have leeway as to which entity to assign to that expression. Thus, for Russell, if there are indefinables of the sort he previously took the cardinal numbers to be in addition to classes of similar classes, then assigning either sort of entity as the reference of numerical expressions will be "adequate to all mathematical uses" of those expressions. The reason Russell

[12]During his Moorean period, Russell regards numbers as properties of plural subjects (see, for example, Russell 1900a, p. 12). This is opposed to the Fregean view Russell later came to accept that "statements of number" are about properties, not objects (see, for example, Russell 1915, pp. 201–2). See Byrd (1987, pp. 65–6) for some discussion of how this change in view is reflected in late additions Russell made to *The Principles of Mathematics*.

chooses to identify cardinal numbers with classes of similar classes rather than indefinable properties common to similar classes is not because he positively denies that there are such indefinables, or because what is "present to his mind" when he understands ordinary numerical terms are classes of similar classes rather than those indefinables, but rather because he is more sure that there are classes of similar classes than that there are indefinable numbers. For Russell whether or not there are such indefinables, there are at least classes of similar classes, and, if they are identified with the cardinal numbers, that is enough to insure that the mathematical statements we take to be true are, in fact, true. But on this view, analysis is no longer a matter of recognizing "the meaning" an expression has by the standard of (Aug)—namely, the entity it stands for—in virtue of its being meaningful at all, "the meaning" with which, by (PoA) together with (Aug), we must be acquainted in order to understand a sentence containing that expression; rather, it is a matter of assigning a precise meaning to an expression that was previously vague—in particular, a precise meaning that insures that the sentences containing that expression have the truth-values we take them as having.

Similarly, in *Principia Mathematica*, Volume II, in defending (Num_{df}), Whitehead and Russell write:

> The chief merits of this definition are (1) that the formal properties which we expect numbers to have result from it; (2) that unless we adopt this definition or some more complicated and practically equivalent definition, it is necessary to regard the cardinal number of a class as an indefinable. Hence the above definition avoids a useless indefinable with its attendant primitive propositions, (Whitehead and Russell 1912, p. 4)

while in the "Preface" to Volume I, they write:

> [W]hen what is defined is (as often occurs) something already familiar, such as cardinal or ordinal numbers, the definition contains an analysis of a common idea, and may therefore express a notable advance. … In such cases, a definition is a "making definite": it gives definiteness to an idea which had previously been more or less vague. (Whitehead and Russell 1910, p. 12)

Previously, Russell held, in accord the Augustinian view of meaning, that the expression "2", for example, had as its meaning a unique entity, simple or complex, and that that it is "the business of philosophy" to ascertain what that entity is. To paraphrase his pre-Peano comments regarding the decision between that absolute and relative theories of magnitude, "it does not lie with us to choose" whether what that expression means is definable or indefinable; "on the contrary, it is the business of philosophy to discover" what it means. Here, however, Russell and Whitehead present us with a choice as to whether to take numbers to be indefinable, or definable as classes of similar classes, or definable in "some more complicated and practically equivalent definition"; and they defend their choice, not on the ground that it reflects what numerical expressions, as ordinarily used, actually mean, or that it reflects what is "present to the mind" when we understand sentences containing numerical expressions, but rather that it enables us to deduce "the formal properties which we expect numbers to have". Compatible with the view that our ordinary notion of cardinal number is "vague", they regard their task not as that of

ascertaining "the meaning" of a numerical expression, as it is ordinarily used, but rather as assigning a precise meaning to such an expression that sustains the claims we wish to make regarding numbers.

Similarly, in *Our Knowledge of the External World*, after introducing (Num_{df}), Russell writes:

> This definition … yields the usual arithmetical properties of numbers. It is applicable equally to finite and infinite numbers, and it does not require the admission of some new and mysterious set of metaphysical entities. (Russell 1915, p. 204)

He then writes in the following paragraph:

> The above definition is sure to produce, at first sight, a feeling of oddity, which is liable to cause a certain dissatisfaction. It defines the number 2, for instance, as the class of all couples, and the number 3 as the class of all triads. This does not *seem* to be what we have hitherto been meaning when we spoke of 2 and 3, though it would be difficult to say *what* we had been meaning. (Ibid.)

And he adds one page later:

> [T]he real desideratum about such a definition as that of number is not that it should represent as nearly as possible the ideas of those who have not gone through the analysis required in order to reach a definition, but that it should give us objects having the requisite properties. Numbers, in fact, must satisfy the formulae of arithmetic; any indubitable set of objects fulfilling this requirement may be called numbers. So far, the simplest set known to fulfill this requirement is the set introduced by the above definition. In comparison with this merit, the question whether the objects to which the definition applies are like or unlike the vague ideas of numbers entertained by those who cannot give a definition, is one of very little importance. (Ibid., p. 205)

Likewise, in *Introduction to Mathematical Philosophy*, he writes:

> [W]hen we come to the actual definition of numbers we cannot avoid what must at first sight seem a paradox …. We naturally think that the class of couples (for example) is something different from the number 2. But there is no doubt about the class of couples: it is indubitable and not difficult to define, whereas the number 2 in any other sense, is a metaphysical entity about which we can never feel sure that it exists or that we have tracked it down. It is therefore more prudent to content ourselves with the class of couples, which we are sure of, than to hunt for a problematical number 2 which must always remain elusive. … In fact, the case of all couples will *be* the number 2, according to our definition. At the expense of a little oddity, this definition secures definiteness and indubitableness; and it is not difficult to prove that numbers so defined have all the properties that we expect numbers to have. (Russell 1919a, p. 18)

Again in these passages, Russell is clear that in identifying cardinal numbers with classes of similar classes, he is not claiming that this captures what we take ourselves to mean by our ordinary statements involving numerical terms; rather, he acknowledges that this account "is sure to produce, at first sight, a feeling of oddity", for what 2 and 3 are on this account "does not *seem* to be what we have hitherto been meaning when we spoke of 2 and 3" and grants, in accord with his earlier view of the cardinal numbers, that we would "naturally" distinguish a cardinal number from a class of similar classes. Again, he does not deny that there are indefinable entities, distinct from classes of similar classes, of the sort he previously

took the cardinal numbers to be; rather, he claims only that we "can never feel sure" that there are such "metaphysical entities", so that is "more prudent" to identify the cardinal numbers with classes of similar classes (which are "indubitable") than with such ("problematic") indefinables. What matters is that we find objects that "satisfy the formulae of arithmetic", and since we can be more sure that there are classes of similar classes than that there are those indefinables, it is "safer" to accept his account of the cardinal numbers.

Again, Russell indicates that there is no fact of the matter as to "*what* we had been meaning" in our ordinary use of arithmetical expressions, since what we entertain when we ordinarily talk of numbers are only "vague ideas". Hence, he indicates that the task of analysis is not, as on his Moorean view, a matter of identifying the entity that is, by the standard of (Aug), "the meaning" a mathematical expression has in its ordinary use, before we undertake the analysis—"the meaning", which, by (PoA) together with (Aug), we must be acquainted with in order to understand a sentence containing that expression, as it is ordinarily used—but is rather, as he indicates in *The Principles of Mathematics*, to find a definition that "is at once precise and adequate to all mathematical uses".

In his 1923 paper on vagueness, Russell holds that a representational system is "precise" when there is a one-one relation between that system and the system it represents, while "a representation is *vague* when the relation of the representing system to the represented system is not one-one but one-many" (Russell 1923, p. 152). For Russell, "a photograph which is so smudged that it might equally represent Brown or Jones or Robinson is vague", and a map is more vague to the extent that "various slightly different courses [of, say, a road or river] are compatible with the representation it gives" (ibid.). Then "[p]assing from representation in general to the kinds of representation that are specially interesting to the logician", he writes:

> [T]he representing system will consist of words, perceptions, thoughts, or something of the kind, and the would–be one–one relation between the representing system and the represented system will be *meaning*. In an accurate language, meaning would be a one-one relation; no word would have two meanings, and no two words would have the same meaning. In actual languages … meaning is one-many. … That is to say, there is not only one object that word means, and not only one possible fact that will verify a proposition. The fact that meaning is a one–many relation is the precise statement of the fact that all language is more or less vague. (Ibid.)

In indicating that in vague language, "meaning" is a "one-many relation", Russell is not claiming that a vague word succeeds in standing for more than one entity; instead, he is indicating that a vague word fails to succeed in picking out, or standing for, one and only one entity as its meaning. That is, a vague word is such that it is compatible with the use of that word that different entities be taken as "the entity" that it stands for, in which case, there is nothing in the use of that word which singles out one of those entities, to the exclusion of the others, as "the entity" it stands for; but this is not to say that one is entitled to take that vague word as standing for all those entities together. Thus, in Russell's example, the smudged photograph that is vague "might equally represent Brown or Jones or Robinson";

but this is not to say that it thereby represents all of them simultaneously. While there is only one person that the photograph can be of, there is nothing in that photograph that determines which person that is.

This account of vagueness is in accord with Russell's post-Peano characterization of numerical expressions as vague. For by indicating there that so far as "all mathematical uses" of numerical expressions goes, we can take those expressions to be referring either to the indefinable that he had previously taken the cardinal numbers to be or to be classes of similar classes, Russell is indicating that there is nothing in our ordinary use of such expressions that determines which (if either) of these sorts of entities those expressions stand for—which is not to say that such expressions stand for both sorts of entities. While analysis will require providing a definite meaning to mathematical terms, there is no assumption that there is a uniquely correct best definite meaning to assign to them.

Post-Peano Analysis: Some Further Examples

Russell not only presents analysis as a matter of proceeding from the "vague" to the "precise" when he defends (Num_{df}); he also characterizes analysis in those sorts of terms when he defends, in his post-Peano writings, definitions of other mathematical concepts. Further, after the publication of PM, he applies this style of analysis to cases that go beyond pure mathematics.

Thus, for example, after the Paris Congress of 1900, but before accepting his logicist definitions of the cardinal numbers, Russell accepted Cantor's and Dedekind's (different[13]) definitions of the infinite as well as Cantor's definition of the continuity; and he defends these definitions in much the same way that he would defend the logicist definitions of the cardinal numbers. In his 1901 paper "Recent Work on the Principles of Mathematics" (later reprinted as "Metaphysics and the Mathematicians"), in introducing the accounts of the infinite given by Cantor and Dedekind, Russell writes:

> [T]hough people had talked glibly about infinity ever since the beginnings of Greek thought, nobody had ever thought of asking, What is infinity? If any philosopher had been asked for a definition of infinity, he might have produced some unintelligible rigmarole, but he would certainly not have been able to give a definition that had any meaning at all. Twenty years ago, roughly speaking, Dedekind and Cantor asked this question, and, what is more remarkable, they answered it. They found, that is to say, a perfectly precise definition of an infinite number or an infinite collection of things. This was the first and perhaps the greatest step. (Russell 1901d, p. 372)

[13]In *The Principles of Mathematics*, Russell (1903, p. 315, see also p. 123) claims that their definitions of finite and infinite numbers "may be easily shown to be equivalent"; however, he later recognizes that their definitions are equivalent only assuming the axiom of choice (or what he and Whitehead call the multiplicative axiom). See (Russell 1903, "Introduction to the Second Edition", pp. viii–ix).

Likewise, in defending Cantor's account of continuity, Russell writes that "Cantor's merit lies, not in meaning what other people mean, but in telling us what he means himself—an almost unique merit, where continuity is concerned" (Russell 1903, p. 353) and also that "as it is certain that people have not in the past associated any precise idea with the word *continuity*, the definition we adopt is, in some degree, arbitrary" (ibid., p. 299). Thus, for Russell, prior to the work of Cantor and Dedekind, the words "infinity" and "continuity" had no definite meaning, so that what Dedekind and Cantor provide is not an account of what was "present to the mind" to those who had previously used those words, but rather precision where there had previously been unclarity. Accordingly, in the "Introduction" to PM, Whitehead and Russell cite Cantor's "definition of the continuum", along with their definitions of the cardinal and ordinal numbers, as cases where definitions "make definite"—that is, give "definiteness to an idea which had previously been more or less vague" (Whitehead and Russell 1910, p. 12).[14]

Again, in his November 1900 draft of *The Principles of Mathematics*, and so before he accepts the logicist definition of cardinal numbers, Russell presents an account of the real numbers, according to which a real number, either rational or irrational, is a class of rational numbers (that he calls a "segment") that is neither null nor co-extensive with the rational numbers, that has no greatest member, and that is such that if y is in S then so is every $x < y$ (Russell 1903, p. 271).[15] Previously, in his pre-Peano draft of *The Principles of Mathematics*, Russell (1899–1900, pp. 114–5) had denied that there are any entities at all that are irrational numbers and had criticized Dedekind and Cantor for accepting "axioms" (including Dedekind's "axiom of continuity") from which it follows that for that for every convergent[16] sequence of rational numbers, there is a number that is the limit of that sequence. Post-Peano, he still rejects the way in which Dedekind and Cantor introduce the irrational numbers—arguing, as he had earlier, that on their theories it "is evidently a sheer assumption" (Russell 1903, p. 281) that for any convergent series of rational numbers there is a number which is the limit of that series.[17] However, once he defines real numbers as segments of rationals, he countenances

[14]Similarly, in discussion following his 1911 presentation of his paper "Analytic Realism", Russell says: "[W]e already have an idea in us of the continuum. But this idea, hitherto vague and unanalyzed, has become precise and analyzed." (Russell 1911b, p. 143). Again, in his *Introduction to Mathematical Philosophy*, Russell writes: "The word 'continuity' had been used for a long time, but had remained without any precise definition until the time of Dedekind and Cantor." (Russell 1919a, p. 100; see also pp. 105–6).

[15]Thus, for Russell (1903, p. 270), the class of all rationals less than 1/2 is the real number 1/2 (which he thereby distinguishes from rational number 1/2); and the class of rational numbers which are such that their squares are less than 2 is the real number $\sqrt{2}$.

[16]Convergent in the sense that the difference between consecutive members of that sequence becomes as small as we like, if they are sufficiently far out in the series (see Russell 1903, p. 281, for this sense of convergence).

[17]Later, he would claim that Dedekind's "method of 'postulating' what we want has many advantages; they are the same as the advantages of theft over honest toil" (Russell 1919a, p. 71).

entities that are irrational numbers. And in defending these definitions of real numbers, he writes:

> [T]here is no logical ground for distinguishing segments of rational numbers from real numbers. If they are to be distinguished, it must be in virtue of some immediate intuition, or of some wholly new axiom, such as, that all series of rationals must have a limit. … My theory, on the contrary, requires no new axiom, for if there are rationals, there must be segments of rationals; and it removes what seems, mathematically, a wholly unnecessary complication, since, if segments will do all that is required of irrationals, it seems superfluous to introduce a new parallel series with precisely the same mathematical properties. I conclude, then, that an irrational actually *is* a segment of rationals which does not have a limit…. (Russell 1903, p. 286, as collated with Byrd 1994, p. 76)

Russell's defense here in November 1900 of his class-theoretic definitions of real numbers anticipates key aspects of his later defense of his class-theoretic definitions of cardinal numbers. In both cases, he contrasts his class-theoretic definitions with definitions according to which the numbers in question are indefinable. In neither case does he deny that there are any indefinable of the sort that others had taken those numbers to be nor does he defend his definitions on the ground that they reflect what had been meant all along by previous uses the defined terms. In both cases, he argues that since his proposed definitions have the benefit of yielding all the required mathematical properties of the numbers to be defined (and hence will "do all that is required"), and since there will be the classes that he uses to define the numbers in question regardless of whether there are any further indefinables, there is no need to assume that those numbers are further indefinables.

After the publication of *Principia Mathematica*, Russell applies his post-Peano conception of analysis to issues that go beyond the philosophy of mathematics. Thus, in 1913, in the first chapter of *Theory of Knowledge*, Russell writes:

> The word "experience", like most of the words expressing fundamental ideas in philosophy, has been imported into the technical vocabulary from the language of daily life, and it retains some of the grime of its outdoor existence in spite of some scrubbing and brushing by impatient philosophers. (Russell 1913, p. 5)

Then, after briefly considering jettisoning the word "experience" altogether, he writes:

> It seems better to persevere in the attempt to analyze and clarify the somewhat vague and muddy ideas commonly called up by the word "experience", since it is not improbable that in this process we may come upon something of fundamental importance to the theory of knowledge, (ibid., p. 6)

remarking more generally:

> A certain difficulty as regards the use of words is unavoidable here, as in all philosophical inquiries. The meanings of common words are vague, fluctuating and ambiguous, like the shadow thrown by a flickering streetlamp on a windy night; yet in the nucleus of this uncertain patch of meaning, we may find some precise concept for which philosophy requires a name. If we choose a new technical term, the connection with ordinary thought is obscured and the clarifying of ordinary thought is retarded; but if we use the common word with a new precise significance, we may seem to run counter to usage, and we may confuse the reader's thoughts by irrelevant associations. It is impossible to lay down a rule for the

> avoidance of these opposite dangers; sometimes it will be well to introduce a new technical term, sometimes it will be better to polish the common word until it becomes suitable for technical purposes. (Ibid.)

The comments Russell makes here echo those he made *The Principles of Mathematics* in acknowledging the "apparent departure from usage" involved, for example, in his definitions of cardinal numbers or in Cantor's definition of continuity. In all these cases, Russell presents analysis as a matter of making precise what was previously vague, so that the task is not to identify what is "present to the mind" of anyone understanding the term as it is ordinarily used, but rather to find a precise meaning to associate with the term—a meaning that is consistent with the central uses of that term or with "nucleus" of the "uncertain patch of meaning" in the term as it is ordinarily used—that is "suitable for technical purposes".

Again, in his 1914 paper "The Relation of Sense-Data to Physics", Russell introduces "the supreme maxim in scientific philosophizing"—namely, "*Wherever possible, logical constructions are to be substituted for inferred entities*" (Russell 1914a, p. 11). As examples of such "logical constructions", he mentions his definitions in *The Principles of Mathematics* of the irrationals and of the cardinal numbers, and his *Principia Mathematica* "no classes" account of discourse apparently about classes. In each case, Russell provides definitions that enable us to prove propositions in the relevant domain without having to hold that there are indefinables of the sort that some (Dedekind and Cantor, in the case of the irrationals; his own earlier self, in the case of the cardinals; and Frege along with his earlier self, in the case of classes) had taken irrationals, cardinal numbers, and classes to be, and without having to accept purported "axioms" (such as Dedekind's "axiom of continuity", or his own earlier "axiom of abstraction", or Frege's Basic Law V) that would guarantee the existence of such indefinables. In none of these cases, does Russell actually deny that there are such indefinables; and in none of them does Russell proceed in accord with the Moorean conception of analysis. For, Russell's central claim is not that his definitions reflect "what we mean" or what is "present to our minds" when we understand sentences containing the defined expressions but rather that given his definitions, the sentences we take to be true in the relevant domains will be true whether or not there are such indefinables.

Accordingly, after mentioning these three cases in which he had "substituted logical constructions" for "inferred entities", Russell continues in "The Relation of Sense-Data to Physics" by writing:

> The method by which the construction proceeds is closely analogous in these and all similar cases. Given a set of propositions nominally dealing with the supposed inferred entities, we observe the properties which are required of the supposed entities in order to make these propositions true. By dint of a little logical ingenuity, we then construct some logical function of less hypothetical entities which has the requisite properties. This constructed function we substitute for the supposed inferred entities, and thereby obtain a new and less doubtful interpretation of the body of propositions in question. This method, so fruitful in the philosophy of mathematics, will be found equally applicable in the philosophy of physics, where, I do not doubt, it would have been applied long ago but for the fact that all who have studied this subject hitherto have been completely ignorant of mathematical logic. (Ibid., p. 12)

The "method" which Russell describes here and seeks to apply to statements of physics is opposed to the Moorean conception of analysis. The task here is not to identify the constituents of the propositions expressed by certain sentences as they are ordinarily used. Rather, it is to find "a new and less doubtful interpretation" of those sentences—"new" in the sense that it does not have to reflect what was "meant" by those understanding those sentences as they are ordinarily used; "less doubtful" in the sense that insures that those sentences in question are true. Russell's purpose in "The Relation of Sense-Data to Physics" is to apply this method, which has been "so fruitful in the philosophy of mathematics", to the philosophy of physics, where he attempts to find an "interpretation" of statements of physics in terms of sense-data that does not assume material objects as "inferred entities".

Similarly, in the final chapter of *Our Knowledge of the External World*, after having discussed in previous chapters not only his theory of cardinal numbers, but also issues concerning continuity, as well as the project of finding an interpretation of statements of physics that treats material objects as "logical constructions", Russell writes:

> The nature of philosophic analysis, as illustrated in our previous lectures, can now be stated in general terms. We start from a body of common knowledge, which constitutes our data. On examination, the data are found to be complex, rather vague, and largely interdependent logically. By analysis we reduce them to propositions which are as nearly as possible simple and precise, and we arrange them in deductive chains, in which a certain number of initial propositions form a logical guarantee for all the rest. … If the work of analysis has been performed completely, [those propositions to which we have reduced the data] will be wholly free from logical redundancy, wholly precise, and as simple as is logically compatible with their leading to the given body of knowledge. (Russell 1915, p. 211)

Like his characterization of "the supreme method of scientific philosophizing", Russell's characterization here of "the nature of philosophic analysis" is a generalization of the conception of analysis that he presents in *The Principles of Mathematics* for accepting, for example, his definitions of the real numbers, the cardinal numbers, and Cantor's definition of continuity. On that view, the task of analysis involves finding precise definitions to assign to terms that had been previously vague, but in such a way as to insure that the sentences containing those terms that we are committed to regarding as true remain true on the precise interpretation proposed. In the terminology Russell uses here, the "initial data" for philosophical analysis is some "given body of knowledge"—sentences that we take to be obviously true, even though the terms in those sentences may be vague. In finding precise definitions for those terms, our task is not to identify what is "present to the mind" of one who understands those terms as they are ordinarily used, since there is nothing definite "present to the mind" of such a person; instead, we our task is to find precise definitions—regardless of how far they may be from how those terms are commonly understood—that help insure the truth of those sentences we ordinarily take to be true involving those terms.

Post-Peano Analysis: Vagueness, Precision, (Aug), and (PoA)

As it is articulated in the passages I have been discussing, Russell's post-Peano, pre-prison practice and conception of analysis makes central use of his distinction between vague and precise language. For Russell, a vague term is such that nothing regarding the use of that term fixes a single entity as its "meaning"; for a precise term, there is a single entity that is rightly regarded as "the meaning" of that term. However, to accept (Aug) is to hold that for an expression to have a meaning is for there to be a single entity that is "the meaning" of that expression. Hence, by the standard of (Aug), an expression that is "vague" in Russell's sense is not meaningful. Moreover, given that there is no single entity that can be regarded as "the meaning" of that expression, understanding that expression cannot be a matter of being acquainted with the entity that is its meaning. In contrast, a precise term is one that meets the standard of (Aug) for being meaningful, so that one is in a position to hold that understanding that term requires being acquainted with the entity that is its meaning.

Hence, prior to going to prison in 1918, Russell has no account as to how, or in what sense, vague language is meaningful or can be understood. Insofar as (Aug) remains his official account of what it is for a term to be meaningful, he should hold that, strictly speaking, vague language is meaningless and so cannot be understood. However, his post-Peano practice and account of analysis assumes that we take certain sentences that include vague terms in them to be true—and it is a condition of success for his account of analysis that we find precise interpretations of such sentences that insure that they are true. Hence, insofar as Russell holds that those sentences that we take to be true at the outset of analysis—those sentences incorporating a "given body of knowledge"—are meaningful and can be understood, he would seem to owe us an account of meaning and understanding according to which vague language can be meaningful and understood.

That he is unconcerned, prior to 1918, with providing such an account of meaning and understanding is, I suggest, because he still holds that there is precise language, which meets the standard of (Aug) for being meaningful, and because he regards vagueness as a defect that it is his task as a philosopher to remedy. Further, for Russell, the positive task that he envisions for philosophy still makes central use of acquaintance and (PoA), albeit not in the way that it does on the Moorean conception of analysis. On the Moorean view, we begin with a sentence S, which we take, by (Aug) to express a definite proposition P, the apprehension of which is required for understanding S. In that case, by (PoA), understanding S requires being acquainted with each constituent of P, so that analysis involves making explicit the entities with which we must already be acquainted in order to understand S.

In contrast, on Russell's post-Peano, pre-prison conception of analysis, we begin with a vague sentence S_v which thereby fails to express any definite proposition but which we take to be obviously true, and we seek to assign to S_v a definite proposition P^* that insures that S_v is true. In apprehending such a proposition, we are, in

virtue of (PoA), thereby acquainted with each of its constituents; hence, we can label each of those constituents and find a precise sentence *S** expressing *P** to replace the vague sentence S_v. While understanding S* requires apprehending P*, and thus, by (PoA), requires being acquainted with each of its constituents, in doing so, we are not thereby making explicit what was "present to the mind" of anyone who understood the original sentence S_v. Here, the task is not to identify what is required for understanding the original vague sentence but rather to replace it with a precise sentence; here, we do not begin with a sentence and then identify the proposition it expresses (or its constituents), but rather independently apprehend a proposition (along with its constituents) and then find a precise sentence to express it. Since the original sentence is vague, understanding it (whatever that may involve) is not a matter of being acquainted with the entities that are the definite meanings of the words in that sentence; on this view, while acquaintance with entities that may be the definite meanings of words in a precise sentence is necessary for analysis, such acquaintance occurs not in understanding the original vague sentence, but rather in order to formulate a precise sentence that may replace the vague one.

Hence, in *The Principles of Mathematics*, when Russell (1903, p. 116) defends his definition of cardinal numbers as classes of similar classes and contrasts it with the view of cardinal numbers as indefinable "predicates" of collections, he writes that "if we can find, by inspection, that there" are such indefinable predicates of collections—that is, if we are acquainted with such predicates—"we may, if we see fit, call this particular class of predicates the class of numbers". However, by adding that "for my part, I do not know whether there is any such class of predicates", he is indicating that he is not acquainted with such indefinables; and by adding that he will accept the definition of cardinal numbers as classes of similar classes "since it is at once precise and adequate to all mathematical uses", he is indicating that while he does not know whether there are any indefinables of the sort he previously took the cardinal numbers to be, he does know—perhaps as a result of acquaintance with them—that there are classes of similar classes. More explicitly, in his 1905 response to Ernest Hobson, Russell presents Hobson as holding that "'the mind' immediately recognizes" indefinable cardinal numbers, and then comments:

> [W]here Dr. Hobson says that "the mind" recognizes such entities, I am unable to agree: if he said "my mind", I should have taken his word for it; but, personally, I do not perceive such entities as cardinal numbers, unless as classes of similar classes. (Russell 1905–06, pp. 78–9)

Thus Russell presents himself as lacking acquaintance with indefinables of the sort he previously took cardinal numbers to be, but as acquainted with classes of similar classes.[18] For Russell, however, while it is thus relevant to the analysis of arithmetic

[18]This despite indicating in the Preface to *The Principles of Mathematics* that, as a result of his "contradiction", he has "failed to perceive any concept requisite for the notion of *class*" (Russell 1903, pp. xv–xvi), and despite introducing, and tentatively, advocating in his reply to Hobson a "no class" theory (see Russell 1905–6, pp. 80–2).

to determine whether or not he is acquainted with indefinables of the sort he previously took the cardinals to be and whether or not he is acquainted with classes of similar classes, what is at issue here is not with identifying the entities with which we must be acquainted in order to understand ordinary sentences involving numerical terms, but rather with finding entities that we "may … see fit … [to] call" the cardinal numbers. That is, we need to find, by acquaintance, entities to serve as constituents of the propositions we are going to express by precise sentences replacing our vague ordinary mathematical sentences.

Russell suggests this sort of view in the course of his correspondence with Victoria Welby. In his post-1918 writings, Russell acknowledges that from early on Welby had encouraged the study of how words have meaning, but that he had not followed her lead. Thus, in his 1959 book *My Philosophical Development*, in discussing his emerging interest in "the problem of the relation of language to facts", Russell writes:

> The problem had been dealt with by various people before I became interested in it. Lady Welby wrote a book about it and F. C. S. Schiller was always urging its importance. But I had thought of language as transparent—that is to say, as a medium which could be employed without paying attention to it. (Russell 1959, p. 14)

Earlier, in his 1926 review of *The Meaning of Meaning*, by Ogden and Richards, Russell writes:

> When, in youth, I learned what was called "philosophy" …, no one ever mentioned to me the question of "meaning". Later, I became acquainted with Lady Welby's work on the subject, but failed to take it seriously. I imagined that logic could be pursued by taking it for granted that symbols were always, so to speak, transparent, and in no way distorted the objects they were supposed to "mean". Purely logical problems have gradually led me further and further from this point of view. (Russell 1926, p. 138)

And the correspondence between Welby and Russell is marked by Welby's repeated attempts to interest Russell in problems concerning how language functions and Russell's repeated indications that he is not concerned with the issues of meaning that she is. Thus, for example, in November 1905, Welby writes:

> I not only learn from students of primitive life and language but realise as part my own deeper experience that while words like 'nothing' are now as you say abbreviations for propositions,[19] the case was originally and now is still in some minds, reversed. Once a word was the only sentence (as before that a sound the only word!) now the sentence—or proposition—is virtually the word. That is why context becomes, in judging the value of a word, so important. … Although I am keenly sensible of my own failure to present the case for 'significs' as an urgent need alike for exact and for poetical or popular thought, I have a great desire to discuss my hopes a little further with you. (Petrilli 2009, pp. 322–3)

[19]She here uses "nothing" as an example of a one-word sentence, because in a previous letter she questions Russell's claim in "On Denoting" that words like "everything", "nothing", and "something" "are not assumed to have any meaning in isolation" (Russell 1905, 416). In particular, she claims that "such words can be used by themselves", as, for example, when in response to the question "What did you give Smith?", one replies "Nothing" (see Petrilli 2009, 321). In reply, Russell wrote that "such words used alone are mere abbreviations for propositions" (ibid., 322).

In reply, Russell writes:

> I think the problem I am studying is not quite the same as yours. I am less concerned with what people do mean than with what things there are that might be meant or would be interesting to be meant. Thus when a single word was the only sentence, I should doubt whether, so far as anything definite was meant, what was meant differed from what we should express by some sentence of many words.[20] I should admit a certain vagueness, which seems to me to be merely ambiguity; but that would be failure to mean any one definite thing, and would not provide a new *thing meant*, as opposite to a new state of mind of the person thinking. But I feel very ignorant in all questions involving the state of mind of a person speaking or thinking. (Ibid., p. 323)

What is relevant here is that Russell makes clear that he is not concerned with how ordinary vague language functions or with "the state of mind" of a person who uses language vaguely. In indicating that he is "less concerned with what people do mean than with things there are that might be meant or would be interesting to be meant", Russell is, I take it, reflecting his concern with replacing ordinary vague language—in which there is a "failure to mean any one definite thing"—by precise language—in which there is a definite "thing meant". For Russell, finding a "new thing meant" will require acquaintance with it—a process that in and of itself does not involve language.

The sense in which Russell, on this view, regards language as a "transparent medium" is not that he regards ordinary, vague language as "transparent", but rather that he regards language as a "medium which *could* be employed without paying attention to it" (my emphasis)—namely, in circumstances in which there is a definite "thing meant". For Russell, that is, once we apprehend a proposition, and thus, by (PoA), are acquainted with each of its constituents, we can then label each of those constituents and formulate a precise sentence in which there is a definite "thing meant". Such a precise sentence will thereby "transparently" represent the proposition it expresses. Thus, the reason Russell held that "logic could be pursued by taking it for granted that symbols were always, so to speak, transparent, and in no way distorted the objects they were supposed to 'mean'", is that he held that the propositions of logic have as their constituents entities—namely, the logical constants—with which we can be acquainted. These are definite "things … that might be meant" if we succeed in becoming acquainted with them; and once acquainted with them, we can then formulate precise, transparent sentences expressing propositions that contain them as constituents.

[20]In his post-1918 writings, Russell (see, for example, 1927, pp. 51ff, 1940, pp. 65ff), discusses one-word sentences at some length. In Chapter "2" of *Word and Object*, Quine (1960, p. 53, note 2) cites Russell's (1940) discussion.

Russell Post-1918: Symbols, Vagueness, and Analysis

Following his prison stay in 1918, Russell's position undergoes a number of changes. What is particularly relevant here is that he not only gives greater prominence to symbols as such but also holds that all symbols are vague. I outline here some aspects of these changes, arguing in particular that accepting some aspects of neutral monism—in particular, no longer countenancing the indefinable relation of acquaintance—plays a central role in leading Russell to regard all language as vague, which in turn requires him to modify his post-Peano conception of analysis.

One change in Russell's view by 1919 is that he has come to hold that the fundamental bearers of truth and falsity ("truth-bearers", for short) essentially contain symbols among their constituents. Initially, following his break with Idealism until 1910, he takes propositions—understood to be unified complexes symbolized by (precise) sentences—to be the fundamental truth-bearers, so that a sentence is true or false only insofar as it expresses a proposition that is true or false. By 1910, in accepting the multiple-relation theory of judgment (MRTJ), Russell no longer countenances such propositions and takes judgments to be the fundamental truth-bearers. On this view, a sentence is true or false only insofar as what it expresses may be judged (or supposed) to be either true or false. Moreover, on the MRTJ, the "objects" of a judgment are the same entities that he had previously regarded as the constituents of a proposition.[21] Hence, on neither of these views is a truth-bearer in the fundamental sense required to contain words, or other symbols, among its constituents; rather, it contains among its constituents the entities (in general, non-linguistic entities) designated by the words in a (precise) sentence corresponding to that truth-bearer.

However, largely as a result of his interaction with Wittgenstein, Russell comes to hold that truth-bearers in the fundamental sense contain symbols among their constituents.[22] In particular, in his post-prison 1919 paper "On Propositions: What They Are and How They Mean", he distinguishes two sorts of "propositions"—"image propositions" and "word propositions" (see Russell 1919b, p. 297). Since images and words are, for Russell, symbols—that is, entities that represent other entities—then (unlike his pre-1910 propositions), these "propositions" are required to have as their constituents symbols themselves and not the entities that the words or images in question symbolize. Moreover, in that paper he holds that these propositions are the fundamental truth-bearers (ibid., Sect. IV).

---

[21] See Footnote 5 above.

[22] Already, by in his pre-prison 1918 lectures "The Philosophy of Logical Atomism", Russell (1918, pp. 165–6) had indicated—as a result, I believe, of in following Wittgenstein in holding that there are no entities that are "logical constants"—that "for the purposes of logic", the sentence should be regarded as the "typical vehicle of truth and falsehood", but he also indicates there that "for the purposes of the theory of knowledge", judgments should be so regarded (where judgments are not regarded as containing symbols, such as images, among their constituents). By 1919, Russell has changed his view of judgment and unequivocally holds that all truth–bearers contain symbols.

Holding that "image-propositions" and "word-propositions" are the fundamental truth-bearers marks a significant change in Russell's view and gives symbols a more prominent place in Russell's philosophy than they had previously. However, taken by itself, this change in his view does not require Russell to give greater prominence to the notion of vagueness, or to hold, in particular, that all language, indeed, all representation, is vague. From Russell's point of view, in considering how "word-propositions" function, he is considering the requirements for a "logically perfect", and hence, precise language. The issue as to how prevalent the phenomenon of vagueness is concerns the relation between ordinary and "logically perfect" language. On the Moorean view, ordinary sentences are "precise" in the sense that they express definite propositions, so that the function of a "logically perfect" language is to provide perspicuous representations of the propositions expressed non-perspicuously. In contrast, as I have argued, on Russell's post-Peano view, ordinary sentences are vague in that they do not express definite propositions, and the task of analysis is to replace a given ordinary vague sentence by a precise sentence couched in a "logically perfect" language that expresses one of the propositions that is consistent with the use of the ordinary vague sentence. So while philosophers may agree as to what the requirements for a "logically perfect" language are, they may disagree as to how much, if any, language is vague in Russell's sense.

In particular, I have argued that in his post-Peano pre-prison writings, Russell indicates that while ordinary sentences are vague, we are capable of replacing them by precise sentences. In his post-1918 writings, in contrast, Russell indicates that all language is vague. Thus, in the second paragraph of his 1923 paper "Vagueness", he writes: "I propose to prove that all language is vague" (Russell 1923, p. 147). And later in that paper, he indicates more generally, that "in practice", no symbols —"words, perceptions, images, or what not"—are "precise" (ibid., p. 150). Hence, in his 1921 Introduction to Wittgenstein's *Tractatus*, while he presents Wittgenstein as "concerned with the conditions for *accurate* Symbolism, i.e. for Symbolism in which a sentence 'means' something quite definite" (Russell 1921b, p. 101), he adds that "[i]n practice, language is always more or less vague, so that what we assert is never quite precise" (ibid.). Similarly, he writes that Wittgenstein is "concerned with the conditions for a logically perfect language—not that any language is logically perfect" (ibid.), so that, for Russell "a logically perfect language" is an "ideal language which we postulate" (ibid.), not a language we can ever realize.[23] I turn now to consider why Russell came to hold following his prison stay that all symbols are

[23]As a number of commentators have discussed (see, for example, Faulkner (2008–9), and others she cites there), in thus presenting Wittgenstein as concerned with "a logically perfect language" as opposed to any actual language, Russell appears to misinterpret Wittgenstein, who writes, for example, in the *Tractatus* that "all the propositions of our everyday language, just as they stand are in perfect logical order" (Wittgenstein 1921, 5.5563). Further, in a 1922 letter to Ogden, Wittgenstein explains this remark by writing:By this I meant to say that the prop[osition]s of our ordinary language are not in any way logically *less correct* or less exact or *more confused* than prop[osition]s written down, say, in Russell[']s symbolism or any other "Begriffsschrift". (Only it is easier for us to gather their logical form when they are expressed in an appropriate symbolism.) (Wittgenstein 1973, 50)

vague; I suggest that it follows, for Russell, from his rejection of *acquaintance*, which, in turn, follows from his coming to accept aspects of neutral monism.

As Russell characterizes it, neutral monism is the view that "both mind and matter are composed of a neutral-stuff which, in isolation, is neither mental nor material" (Russell 1921a, p. 25). While he comes to hold, by 1919, that sensations are "neutral entities" that "belong[ ] equally to psychology and to physics", he holds that some entities—namely, images—"belong only to the mental world", while there may be other entities that "belong only to the physical world" (ibid.). However, he agrees with "neutral monists", that minds are no longer to be regarded as among the ultimate constituents of the universe but are rather to be "constructed", in which case he can no longer accept his former view of acquaintance. As he writes in "On Propositions":

> I have to confess that the theory which analyses a presentation into act an object no longer satisfies me. The act, or subject, … seems to serve the same sort of purpose as is served by points and instants, by numbers and particles and the rest of the apparatus of mathematics. All these things have to be *constructed*, not postulated: they are not the stuff of the world, but assemblages which it is convenient to be able to designate as if they were single things. … It seems to me imperative, therefore, to construct a theory of presentation … which makes no use of the "subject", or of an "act" as a constituent of a presentation. (Russell 1919b, p. 294)

In 1911, Russell had written: "[T]o say that *S* has acquaintance with *O* is essentially the same thing as to say that *O* is presented to *S*" (Russell 1911a, p. 148). Hence, by indicating here that having come to regard "the subject" as a "construction", he can no longer accept his former theory of "presentation", he is thereby indicating that he can no longer accept his former theory of acquaintance. Given that there are no ultimate constituents of the world that are minds, there can be no indefinable relation of acquaintance that relates minds to other entities. Hence, he will have to provide a different account of what is involved when we speak, for example, of "objects" of thought or of perception.

I have argued that it is in virtue of his continuing commitment, during his post-Peano pre-prison period, to the indefinable relation of acquaintance that Russell is able to hold that we are capable of replacing vague with precise language. For Russell, while ordinary language may be vague, we can, by means of acquaintance, have a direct, unmediated cognitive relation to simple extra-mental entities, and once we are acquainted with such entities, we can then label them and formulate precise "logically perfect" sentences, in which each word stands for a simple entity with which we are acquainted. On this view, acquaintance occurs independently of language, or, indeed, of any sort of representation or symbol; precise language thus follows acts of acquaintance with entities that can then serve as the meanings of precise symbols. Acquaintance provides the means by which we can anchor precise language.

(Footnote 23 continued)

From this remark, at any rate, it would appear that Wittgenstein's early conception of analysis, unlike Russell's, is in accord with the Moorean view.

Accordingly, once Russell rejects this relation of acquaintance, he holds that all language is vague. First, instead of regarding acquaintance as a primitive, unmediated relation between thinking subjects and extra-mental "objects", Russell now writes in *The Analysis of Mind*:

> The reference of thoughts to objects is not, I believe, the simple direct essential thing that Brentano and Meinong represent it as being. It seems to me to be derivative, and to consist largely in *beliefs*: beliefs that what constitutes the thought is connected with various other elements which together make up the object. You have, say, an image of St. Paul's, or merely the word "St. Paul's" in your head. You believe, however vaguely and dimly, that this is connected with what you would see if you went to St. Paul's, or what you would feel if you touched its walls; it is further connected with what other people see and feel, with services and the Dean and Chapter and Sir Christopher Wren. These things are not mere thoughts of yours, but your thought stands in a relation to them of which you are more or less aware. The awareness of this relation is a further thought, and constitutes your feeling that the original thought had an "object". … Thus the whole question of the relation of mental occurrences to objects grows very complicated, and cannot be settled by regarding reference to objects as of the essence of thoughts. (Russell 1921a, pp. 18–19)

In this passage, Russell is not merely criticizing the view of Brentano and Meinong; he is criticizing his own former view of acquaintance. Whereas previously he took acquaintance to be "simple [and] direct", more primitive than belief, and independent of representation, he now regards "the relation of mental occurrence to objects" as "very complicated" and as a "derivative" phenomenon that "consist[s] largely in beliefs" and that involves symbols, such as images or words. And it is understandable that having rejected the primitive relation of acquaintance by which to ground precise language, and holding instead that "the relation of mental occurrences to objects" involves beliefs we hold, "however vaguely and dimly", Russell will be led to hold that no symbol we use—no image or word—will succeed in "meaning" something "quite definite".

By rejecting acquaintance, Russell lacks the means by which we can secure a precise or definite meaning for a word; and by rejecting acquaintance, he can no longer hold—as he held both on his early view of propositions and on the MRTJ—that understanding a (precise) sentence requires being acquainted with the entities designated by the words in that sentence. Accordingly, in *The Analysis of Mind*, he presents a different account of understanding and meaning, an account that he immediately connects with the view that all language is vague. As he writes, in a passage, most of which I quoted above:

> It is not necessary, in order that a man should "understand" a word, that he should "know what it means," in the sense of being able to say "this word means so-and-so." Understanding words does not consist in knowing their dictionary definitions, or in being able to specify the objects to which they are appropriate. … Understanding language is more like understanding cricket: it is a matter of habits, acquired in oneself and rightly presumed in others. To say that a word has a meaning is not to say that those who use the word correctly have ever thought out what the meaning is: the use of the word comes first, and the meaning is to be distilled out of it by observation and analysis. Moreover, the meaning of a word is not absolutely definite: there is always a greater or less degree of vagueness. The meaning is an area, like a target: it may have a bull's eye, but the outlying parts of the target are still more or less within the meaning, in a gradually diminishing

> degree as we travel further from the bull's eye. As language grows more precise, there is less and less of the target outside the bull's eye, and the bull's eye itself grows smaller and smaller; but the bull's eye never shrinks to a point, and there is always a doubtful region, however small, surrounding it. (Ibid., pp. 197–8)

Above I argued that although the view that much of language is vague is central to Russell's post-Peano pre-prison view of analysis as proceeding from the vague to the precise, he did, not during that period, present an account as to how vague language can be meaningful and understood—perhaps because he regarded vague language as a defect that it is the task of philosophy to remedy. However, once he rejects his former view of acquaintance, he no longer has the means to remedy that defect. Further, given that he now regards language, and, more generally, symbols, as playing a central role as truth-bearers, it is incumbent on him to provide and an account of meaning and understanding that does not rely on acquaintance. The account he accepts following his prison stay, and heavily influenced by his reading of Watson, is that understanding is a matter of using a word in "suitable circumstances" and reacting to words with "suitable behavior", where what is "suitable" is "a matter of habits, acquired in oneself and rightly presumed in others", and that "meaning is to be distilled out of [use] by observation and analysis". Moreover, Russell suggests that one consequence of this view is that "the meaning of a word" is never "absolutely definite": given that "the meaning" of a word is to be "distilled out" of its "use", and given that the use will never fix a precise meaning, then all language will be vague to at least some extent. On his post-Peano, pre-prison conception of analysis, Russell agreed, in effect, that our mere uses of a word does not determine a unique meaning for that word—so that, for example, our use of arithmetical statements does not thereby provide a unique meaning for arithmetical words—but he held also that acquaintance with extra-mental entities provides us with the means to assign precise meanings to words in a logically perfect language. Post-prison, Russell retains the view that use does not fix meaning uniquely and no longer accepts a notion of acquaintance capable of providing us with precise meanings.

Given that Russell no longer holds that any language is precise, he can no longer present analysis as a process that moves from the vague to the precise; rather, he must present it as a process that moves from the more to the less vague, without ever attaining absolute precision. Thus, for example, in the first chapter of his 1927 book *Philosophy*, he writes:

> Philosophy arises from an unusually obstinate attempt to arrive at real knowledge. What passes for knowledge in ordinary life suffers from three defects: it is cocksure, vague, and self-contradictory. The first step towards philosophy consists in becoming aware of these defects, not in order to rest content with a lazy skepticism, but in order to substitute an amended kind of knowledge which shall be tentative, precise, and self-consistent, (Russell 1927, pp. 1–2)

adding two paragraphs later:

> I mentioned a moment ago three defects in common beliefs, namely, that they are cocksure, vague, and self-contradictory. It is the business of philosophy to correct these defects so far as it can, without throwing over knowledge altogether. … All these, of course, are a matter of degree. Vagueness, in particular, belongs in some degree, to all human thinking; we can diminish it indefinitely, but we can never abolish it wholly. Philosophy, accordingly, is a continuing activity, not something in which we can achieve final perfection once and for all. (Ibid., p. 3)

Again, in his 1940 book *An Inquiry into Meaning and Truth*, he writes:

> I should say that inquiry begins, as a rule, with an assertion that is vague and complex, but replaces it, when it can, by a number of separate assertions each of which is less vague and less complex than the original. (Russell 1940, p. 320)

And in his 1948 book, *Human Knowledge: Its Scope and Limits*, in a chapter entitled "Interpretation", he writes:

> It often happens that we have what seems adequate reason to believe in the truth of some formula expressed in mathematical symbols, although we are not in a position to give a clear definition of the symbols. It happens also, in other cases, that we can give a number of different meanings to the symbols, all of which will make the formula true. In the former case we lack even one definite interpretation of our formula, whereas in the latter we have many. This situation, which may seem odd, arises in pure mathematics and in mathematical physics; it arises even in interpreting common sense statements such as "My room contains three tables and four chairs." It will thus appear that there is a large class of statements, concerning each of which in some sense we are more certain of its truth than of its meaning. "Interpretation" is concerned with such statements; it consists in finding as precise a meaning as possible for a statement of this sort, or, sometimes, in finding a whole system of possible meanings, (Russell 1948, pp. 235–6)

writing elsewhere in that book:

> Philosophy, like science, should realize that, while complete precision is impossible, techniques can be invented which gradually diminish the area of vagueness or uncertainty. (Ibid., p. 147)

In all these passages, Russell presents the same sort of account of analysis that I have argued above he introduces in his post-Peano writings—an account he uses first to justify his technical definitions of mathematical concepts, then generalizes in his writings after *Principia Mathematica* to apply more broadly. The only difference is that whereas in his pre-prison writings he describes analysis as proceeding from the vague to the precise, in these post-prison writings he characterizes it as proceeding from the vague to the less vague or more precise. Lacking acquaintance as the means by which to secure precise language, he now presents analysis as a "continuing activity", not as "something in which we can achieve final perfection once and for all".

There are, thus, significant differences between the role that language, and in particular vague language, plays in Russell's overall philosophy, and, more specifically, in his conception of analysis, before and after his prison stay in 1918. Before his prison stay, Russell attempts to retain, for at least some cases, the view that fundamental truth-bearers are judgment-complexes that do not contain symbols among their constituents; after his prison stay, he holds that the fundamental truth-bearers are "word-propositions" or "image-propositions", complexes that contain symbols among their constituents. Before his prison stay, Russell retains his early notion of acquaintance, and in doing so has the means to account for how we can give words precise meaning; after his prison stay, he rejects that notion of acquaintance and presents an account of meaning according to which all words are vague. Hence, before his prison stay, while Russell acknowledges the phenomenon of vague language, he regards it as a flaw that can be corrected, expresses no interest in considering how, or in what sense, vague language can be meaningful and understood, and, accordingly, presents no account of meaning or understanding that takes into account the phenomenon of vague language. After his prison stay, in regarding language—and, more generally, symbolism—as not only playing a central role in the theory of truth but also as unavoidably vague, Russell explicitly theorizes about the nature of language, and does so in such a way as to explain how vague language can be meaningful and understood.

However, despite these differences between Russell's view pre- and post-prison, on his post-Peano conception of analysis before and after his prison stay—and as opposed to the Moorean conception of analysis—what initiates analysis is vague language. For that reason, both before and after his prison stay, and as opposed to the Moorean conception of analysis, Russell rejects the view of analysis as making explicit what is "present to the mind" of anyone who understood the original sentence in question, prior to analysis, and holds, instead, that analysis is a matter of replacing a vague sentence that we are committed to regarding as true by a less vague (if not absolutely precise) sentence by which we can "interpret" the original sentence as being true. I argue now that by failing to appreciate this aspect of Russell's post-Peano view of analysis, Hylton is led not only to find certain aspects of Russell's post-Peano practice of analysis puzzling but also to fail to recognize certain fundamental similarities between Quine's view of analysis and that of the post-Peano Russell.

Hylton, Quine, and Russell's Post-Peano Analysis

In his 2007 paper "'On Denoting' and the Idea of a Logically Perfect Language", Hylton begins by focusing on the following passage from Russell's pre-prison 1918 lectures on logical atomism:

> In a logically perfect language the words in a proposition would correspond one by one with the components of the corresponding fact, with the exception of such words as "or", "not", "if", "then", which have a different function. In a logically perfect language, there will be one word and no more for every simple object, and everything that is not simple will be expressed by a combination of words …. The language which is set forth in *Principia Mathematica* is intended to be a language of that sort. … Actual languages are not logically perfect in this sense, and they cannot possibly be, if they are to serve the purposes of daily life. (Russell 1918, p. 176; quoted by Hylton 2007, p. 91)

In the course of discussing this passage, Hylton writes:

> A … point, which does not emerge explicitly in the passage quoted, is that the fully analysed sentence corresponds to the thought which is expressed by the ordinary, unanalysed sentence. In *Problems of Philosophy*, speaking of definite descriptions, Russell says: 'the thought in the mind of a person using a proper name correctly can generally only be expressed explicitly if we replace the name by a description' …. So the fully analysed sentence has a structural correspondence with something which is psychologically real. In other places he is not so explicit, but I think his underlying view must be the same. Russell is committed to the view that the fully analysed sentence has a structural correspondence with something which is expressed both by that sentence and by its unanalysed version. … Our ordinary sentences imperfectly express our thoughts. The complete analysis of a sentence, if it could be carried out, would yield a sentence whose structure corresponds to that of the thought which the original sentence imperfectly expresses …. (Hylton 2007, p. 93)

Here Hylton attributes to the Russell the Moorean conception of analysis—a conception according to which the task of analysis is to represent in a perspicuous way what is "present to the mind" of the person who understands an ordinary, unanalyzed sentence. On this view, the difference between an ordinary, unanalyzed sentence and a corresponding fully analyzed one is not what thought they express, but rather how they express that thought: while the ordinary sentence represents that thought "imperfectly"—in a way that does not reveal the ultimate structure of what is expressed—the fully analyzed sentence expresses exactly the same thought, but in such a way that the structure of that sentence corresponds to the structure of the thought expressed.

As I have indicated, I agree that there are at least some passages in his post-Peano writings in which Russell presents that view of analysis, not least in passages in which he is defending (PoA); and the passage from *The Problems of Philosophy* that Hylton here cites is from the chapter in which Russell defends (PoA).[24] However, I have also argued that on the view of analysis that derives from his post-Peano analyses in mathematics, Russell makes clear that he is not adhering to this conception of analysis. Thus, he never presents his definitions of cardinal numbers, or of irrational numbers, or, eventually, of matter or mind, as

[24]It might be noted, however, that two paragraphs following the passage that Hylton quotes, Russell writes: "When we, who did not know Bismarck, make a judgment about him, the description in our minds will probably be some more or less vague mass of historical knowledge …." (Russell 1912, p. 55). For some further discussion of what sort of "analysis" Russell intends his theory of descriptions as providing, see Footnote 27 below.

perspicuously representing what is "present to the mind" of—or what is "psychologically real" to—ordinary users the expressions in question.

Hylton acknowledges that in the passage regarding "a logically perfect language" that he quotes from "The Philosophy of Logical Atomism", Russell does not "explicitly" claim that such a language captures what is "psychologically real" to users of ordinary, logically imperfect languages; nevertheless, he claims that Russell is "committed" to this view. However, early in the first lecture in "The Philosophy of Logical Atomism", Russell characterizes his "method of analysis" as "pass[ing] from the vague to the precise", writing in particular:

> It is a rather curious fact in philosophy that the data which are undeniable to start with are always rather vague and ambiguous. You can, for instance, say: "There are a number of people in this room at this moment." That is obviously in some sense undeniable. But when you come to try and define what this room is, and what it is for a person to be in a room, and how you are going to distinguish one person from another, and so forth, you find that what you have said is most fearfully vague and that you really do not know what you meant. That is a rather singular fact, that every thing you are really sure of, right off is something you do not know the meaning of, and the moment you get a precise statement you will not be sure whether it is true or false, at least right off. The process of sound philosophizing, to my mind, consists mainly in passing from those obvious, vague, ambiguous things, that we feel quite sure of, to something precise, clear, definite, which by reflection and analysis we find is involved in the vague thing that we start from, and is, so to speak, the real truth of which that vague thing is a sort of shadow. … Everything is vague to a degree you do not realize till you have tried to make it precise, and everything precise is so remote from everything that we normally think, that you cannot for a moment suppose that is what we really mean when we say what we think. (Russell 1918, pp. 161–2)

Here, as in other writings I have cited above from both pre-1918 and post-1918 writings, Russell presents the view of analysis which emerges from his post-Peano "analyses" of mathematical concepts, but which is opposed to Moorean views of analysis and meaning.[25] Thus, in claiming that although "the data" with which we begin in philosophy are claims we take to be obviously true, those claims themselves are "fearfully vague", so that "you really do not know what you meant" in making such claims, Russell is characterizing his post-Peano procedure of beginning with statements in a given domain that we take to be true and then attempting to assign definite meanings to the expressions in those statements that, while insuring that those statements are true, are not intended to reflect "what we really mean"—or anything that was "psychologically real"—when we originally made those statements. Here the relation between an ordinary unanalyzed sentence and an analyzed version of it is not that between non-perspicuous and perspective expressions of the same proposition but rather between a vague representation, which does not determine a unique proposition as "the" proposition expressed, and a precise representation, which determines a single proposition as the proposition expressed, that has the same truth-value we assign to the vague representation. (And here, as in other pre-prison writings, and consistent with his still employing the

[25]See also Russell (1959, p. 133), where he characterizes this view of his "method" as "my strongest and most unshakable prejudice as regards the methods of philosophical investigation".

notion of acquaintance in "The Philosophy of Logical Atomism",[26] Russell indicates that the outcome of analysis will be precise statements.)

Further, in failing to distinguish Russell's Moorean conception of analysis from his post-Peano analyses in mathematics, Hylton finds it puzzling that in discussing a number of his analyses in which he avoids commitment to there being entities of a given sort, Russell nevertheless allows that, despite his analysis, there may, in fact, be entities of the sort that he has just avoided countenancing. As I have discussed, his post-Peano account of cardinal numbers is such a case: while his definitions of the cardinals allow Russell to avoid holding that there are indefinable numbers of the sort he had admitted on his Moorean absolute theory of number, in introducing these definitions, Russell does not deny that there are such indefinables, and claims only that whether or not there are such indefinables, we do not need to assume them in order for our arithmetical statements to be true. While Hylton does not mention this case, he discusses similar claims Russell makes regarding classes on the "no class" theory he develops in *Principia Mathematica*. Thus, after writing that on Russell's analysis "the truth of sentences using [class] symbols is … explained without supposing that there are classes, and our ability to use those symbols, even though we have no epistemic access to classes, is explained", he continues:

> But he then sometimes goes on to say that classes may exist anyway, independent of our definitions. … How he thinks it can even make sense for him to say this, given his other commitments, is very far from clear. (Hylton 2007, p. 105, note 1)

On the Moorean conception of analysis, words have a definite meaning and understanding words requires that we have that meaning "present to the mind". Hence, our numerical expressions either refer to indefinables or they don't; our class symbols either refer to classes or they don't. But however they contribute to the meaning of the sentences in which they occur, that meaning is "present to the mind" of one who understands those sentences. Hence, if the analysis is correct, it reveals what we really mean when we use class talk, in which case it makes no sense to defend an analysis of class talk which does not assume that there really are any entities that are classes, but then go on to ask whether there really are classes. For what, then, is the meaning of class-talk in this question? In contrast, on the view that ordinary class talk is vague, it is coherent to suppose that there are different interpretations of class-talk all of which yield the truth-values we want to assign to sentences involving class symbols, in which case it is coherent to raise the question as to whether there are other entities besides those we have used to interpret class-talk that may be taken as the meaning of class symbols. And this is Russell's consistent pattern when he avoids a given sort of indefinable—whether it be irrationals, as Dedekind and Cantor understood them, or cardinal numbers, as he had originally understood them, or classes, or matter, or minds. In each case, when he presents a "logical construction" that enables us to avoid countenancing certain

[26]Thus, for example, he writes: "All analysis is only possible in regard to what is complex, and it always depends, in the last analysis, upon direct acquaintance with the objects which are the meanings of certain simple symbols" (Russell 1918, p. 173).

"inferred" entities, he does not actually claim that there are no such indefinables, only that we do not need to assume such indefinables in order to interpret the relevant sentences as true.[27]

Given that assuming that Russell adheres consistently to a Moorean conception of analysis requires finding aspects of Russell's post-Peano practice of analysis problematic, it is not surprising that Moore himself would object to Russell's post-Peano definitions. Accordingly, in an unpublished review of PoM, written apparently some time in 1905, Moore writes:

> But [Russell's] definition in logical terms of number 'one' is by no means simple … It is not plain that what we think to be true of the penny, when we think it is but *one*, is no less than that it is a member of the class of [one-membered] classes …: it is not plain that this is a correct *analysis* of what we think. That it is *equivalent* to what we think, in the sense that anything whatever which has the property which we mean by 'one' is also a member of this class of classes, and that anything whatever which is a member of this class of classes also has the property which we mean by 'one', there is, indeed, no doubt whatever. But Mr. Russell admits the possibility that it is *only* equivalent—that, possibly, all the members of this class of classes have in common some *other* property, beside the fact that they belong to this class—some other property, which belongs to all of them and only to them, and which may be what we generally mean when we speak of the number 'one'. Mr. Russell, indeed boldly asserts his doubt whether there is any such other property; and there is much to be said for his view. But what I wish now to point out is the consequences which follow

[27]See, for example, (Russell 1918, p. 237), where he makes the point with regard to matter, and (Russell 1919b, p. 294), where he makes the point with regard to minds. In contrast, in *Principia Mathematica*, Whitehead and Russell distinguish Russell's theory of descriptions in this regard from their "no class" theory, writing that in "[t]he case of descriptions it was possible to *prove* that they are incomplete symbols" but that "in the case of classes, we do not know of any equally definite proof" and that "it is not necessary … for our purposes to assert dogmatically that there are no such things as classes" (Whitehead and Russell 1910, p. 75). Again, while Russell consistently presents any of his "logical constructions" as providing one, but not the only possible, interpretation of the discourse in question, in his (1905) he takes himself to present decisive arguments against other proposed analyses of propositions expressed by sentences containing definite descriptions. However, in "On Denoting", Russell acknowledges that his "interpretation" of propositions expressed by sentences of the form "The *F* is *G*" "may seem … somewhat incredible" (Russell 1905, p. 417), thus suggesting that he does not regard his "interpretation" as reflecting what is "present to the mind" of one who understands such a sentence. And in his 1957 reply to Strawson, which he reprints in *My Philosophical Development*, Russell writes:I … am persuaded

> that common speech is full of vagueness and inaccuracy, and that any attempt to be requires modification of common speech both as regards vocabulary and as regards syntax. … My theory of descriptions was never intended as an analysis of the state of mind of those who utter sentences containing descriptions. … I was concerned to find a more accurate and analysed thought to replace the somewhat confused thoughts which most people at most times have in their heads, (Russell 1959, pp. 241–243)

Thus, while there are some significant differences between the way Russell presents his theory of definite descriptions and how he presents his "logical constructions", here, at least, he presents it as conforming to his post-Peano model of analysis of replacing the "vague" and "confused" with something more "precise and accurate". For some further discussion of the issues raised here, see Szabo (2005, Sect. 2) and Kripke (2005, 1107, note 28).

> from the mere possibility that there is such another concept, meant by 'one'. … When Mr. Russell asserts that 1 + 1 = 2 can be deduced from logical principles, his assertion only applies to the proposition in which the concept dealt with is 'the class of classes, of which each etc. etc'; it is only *this* proposition which he shows to be deducible from logical principles. If it be true that there is also *another* concept denoted by the word 'one', then the proposition that 1 + 1 = 2, understood as asserting a universal connection between this *other* concept and some others, *cannot* be deduced from logical principles alone. … Unless, therefore, it can be shown that the concepts dealt with in those propositions, which can be deduced from logical principles, are the very ones which occur in the proposition 1 + 1 = 2, as ordinarily understood, then it must be admitted *either* that the proposition 1 + 1 = 2, as ordinarily understood, is not a proposition of pure mathematics *or* that Mr. Russell's [logicism] does not … apply to all propositions of pure mathematics. (Moore 1905, pp. 8–10)

Here, Moore, in effect, assumes his conception of analysis in criticizing Russell's definitions of the cardinal numbers. Thus he questions whether Russell's definition is "a correct *analysis* of what we think" when we make such a claim as "There is one penny in my pocket" or of "what we generally mean when we speak of the number 'one'", or whether he provides a correct account of the "concepts" which "occur in the proposition 1 + 1 = 2 as ordinarily understood".[28] Like Hylton, he is assuming that Russell is committed to capturing what is "psychologically real" when we understand, for example, "1 + 1 = 2". And since he finds it implausible to suppose that Russell's definition of the number one—a definition "which is by no means simple"—does capture "what we generally mean when we speak of the number 'one'", then Moore indicates, in accord with his conception of analysis, that Russell's definition cannot be correct.[29]

Further, like Hylton, Moore finds it problematic that Russell is agnostic as to whether there are, in addition to classes of similar classes, indefinable properties of the sort Russell previously took numbers to be. For, given Moore's view, in accord with (Aug), that there is a single entity, which is "the meaning" of the expression "one", as it is ordinarily used, then it is not up to Russell to choose to define numbers as classes of similar classes if numbers are really indefinables. Hence, for Moore, if the proposition "1 + 1 = 2 as ordinarily understood" contains among its constituents the indefinables Russell previously took numbers to be, then that

[28]See, similarly Husserl, who assumes what I have called the Moorean conception of analysis in criticizing an equivalence class definition of faintness of tone:'What we mean' is surely our sense, and can one say even for an instant that the sense of the proposition 'This tone is faint' is the same as the sense of the proposition 'This tone belongs to a group (of whatever sort) of similars'?.… Naturally the utterances 'A tone is faint' and 'A tone belongs to the sum total of objects alike in their faintness' are semantically equivalent, but equivalence is not identity. (Husserl 1900–1, pp. 303–4).

[29]Likewise, Moore criticizes the two definitions of the infinite that Russell (following Dedekind and Cantor) provides in *The Principles of Mathematics* in favor of a definition that he claims "is far more in accordance with the ordinary use of the word 'infinity'"(Moore 1905, pp. 30–1) than those presented by Russell.

proposition does not consist wholly of constituents definable in purely logical terms, in which case if the proposition "1 + 1 = 2 as ordinarily understood" is a proposition of pure mathematics, then Russell's logicism is false. For Moore, as for the pre-Peano Russell, we are not free to introduce definitions as a matter of "convenience"; they have to be answerable to what our words "as ordinarily understood" actually mean. Whereas the post-Peano Russell takes our ordinary use of numerical expressions as vague, in which case he allows himself to introduce precise definitions "adequate to all mathematical uses" of such expressions that enable him to sustain his logicism, for Moore, Russell is not free to introduce his definitions of cardinal numbers simply because they facilitate his logicism; if they do not correspond to "what we generally mean" when we use numerical expressions, then so much the worse for logicism.

Finally, in assuming that the "pre-linguistic" Russell is committed to the Moorean conception of analysis, Hylton is led to draw a sharp contrast between Russell's view of analysis and Quine's. In particular, Hylton writes:

> Quine speaks of the definition of ordered pair, either by the method of Weiner or by that of Kuratowski, as a 'a philosophical paradigm'…. Right away we see a difference between Quine and Russell. Wiener's method is not the same as Kuratowski's. … But then which method is correct? Which most closely reflects the underlying structure of propositions in which reference is made to ordered pairs? For Quine, unlike Russell, these are misleading questions, better rejected than answered. (Hylton 1996, pp. 47–8)

As should be clear, the contrast Hylton draws here is between Quine and the Moorean Russell, not Quine and the post-Peano Russell. For the Moorean Russell, each word has a definite meaning, and it is up to the philosopher to "discover" what that meaning is, so that between two different analyses of a given expression at most one of them will be correct, at most one of them will perspicuously represent what is "present to the mind" of one who understands a sentence containing that expression. For the post-Peano Russell, in contrast, our use of mathematical expressions is vague, so that prior to analysis, there is no one definite meaning the analysis has to answer to, in which case there may be different equally adequate ways to make precise the original vague expression. While the Moorean Russell is required to ask of two different proposed analyses, "Which is correct?", the post-Peano Russell, like Quine, regards this as a misleading question resulting from wrongly assuming a determinacy of meaning in the original expression prior to analysis.

Hylton then goes on to quote from *Word and Object*, where, in discussing the set-theoretical definition of "ordered-pair", Quine writes:

> This construction [of the ordered pair] is paradigmatic of what we are most typically up to when in a philosophical spirit we offer an "analysis" or "explication" of some hitherto inadequately formulated "idea" or expression. We do not claim synonymy. We do not claim to make clear and explicit what the users of the unclear expression had unconsciously in mind all along. We do not expose hidden meanings, as the words 'analysis' and 'explication' would suggest; we supply lacks. We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions. (Quine 1960, pp. 258–9; quoted by Hylton 1996, p. 48)

Hylton comments:

> Quine's appeal to the definitions of Wiener and Kuratowski clearly represents a continuation of a trend that Russell, along with Frege, began: the use of technical methods in philosophy. What is striking, however, from the present point of view, is how the technical methods stand aloof from the philosophical disagreement. … The technical method is the same, yet the philosophical purpose, the philosophical gloss, is about as different as could be. From Quine's point of view, his version of, or substitute for, philosophical analysis is a way of preserving the insights of Russell and others without their excess metaphysical baggage. From the point of view of Russell, and indeed of many current authors, Quine has thrown out the baby with the bathwater. (Hylton 1996, p. 48)

As should be clear, however, while Quine's characterization of the method of analysis revealed by the construction of the ordered pair is fundamentally opposed, in the way Hylton suggests, to Russell's Moorean conception of analysis, it is in accord with Russell's post-Peano view of analysis. Like Russell, Quine denies that there was one definite meaning that that phrase had prior to the analysis. Hence, like Russell, Quine does not present himself as articulating "the" meaning that the phrase to be analyzed had prior to the analysis or the meaning that the users of that phrase "had unconsciously in mind all along". Like Russell, Quine presents the analysis as a matter of replacing an unclear expression which fulfilled certain functions by a clear expression that fulfills those same functions. That is to say, in his characterization in this passage of analysis, Quine is not the "heretical follower" of Russell, but rather his faithful student. Moreover, it is not the Russell of the 1920s that Quine is here emulating, but rather the post-Peano Russell of PoM. The renegade here is not Quine, but Russell, who by introducing "technical methods" into philosophy, undermined his own post-Idealist Moorean views.

There are, I recognize, significant differences between Quine and the pre-prison Russell. For by countenancing the indefinable relation of acquaintance, the pre-prison Russell has an account as to how language can be made precise that Quine, along with the post-prison Russell, would reject. Further, while the pre-prison Russell presents no account as to how vague language can be meaningful or understood, and presents only an account as to how precise language is meaningful and can be understood that is in accord with (Aug) and (PoA), Quine, like the post-prison Russell, presents a behaviorist-inspired account of language that is opposed to (Aug) and (PoA) and that explains how vague language can be meaningful and can be understood. However, it is in his post-Peano writings, almost two decades before his prison stay, that Russell introduces the notion of vagueness in order to justify his technical definitions of such mathematical concepts as the cardinal numbers, the irrationals, continuity, and infinity. Thus, he introduces the view that there is a kind of indeterminacy in our ordinary use of numerical terms, so that neither they nor sentences in which they occur have a determinate meaning, a kind of indeterminacy that is central to the view of analysis he presents in the Introduction to *The Principles of Mathematics* as a matter of "giv[ing]

precision to … notion[s] which had hitherto been more or less vague".[30] Whereas Quine presents this sort of view of analysis (albeit, without adhering to the conception of precision that Russell accepted until his 1918 prison stay) in Chapter "5" of *Word and Object* as following from his Chapter "2" rejection of synonymy and determinate meaning, Russell introduces examples of indeterminate language and makes them central to his view of analysis almost two decades before he presents an account of language from which it follows that vague language can be meaningful and understood.[31]

References

Baldwin, T. (2003). From knowledge by acquaintance to knowledge by causation. In N. Griffin (Ed.), *The Cambridge Companion to Bertrand Russell* (pp. 420–448). Cambridge: University Press.

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

Byrd, M. (1987). Part II of *The Principles of Mathematics*. *Russell*, n.s. 7, pp. 60–70.

Byrd, M. (1994). Part V of *The Principles of Mathematics*. *Russell*, n.s. 14, pp. 47–86.

Byrd, M. (1996–7). Parts III–IV of *The Principles of Mathematics*. *Russell*, n.s. 16, pp. 145–68.

Dreben, B. (1996). Quine and Wittgenstein: The odd couple. In R. Arrington & H-J. Glock (Eds.), *Wittgenstein and Quine* (pp. 39–61). London and New York: Routledge.

Faulkner, N. (2003). Russell and Vagueness. *Russell*, n.s. 23, pp. 43–63.

Faulkner, N. (2008–9). Russell's Misunderstanding of the *Tractatus* on Ordinary Language. *Russell*, n.s. 28, pp. 143–63.

Griffin, N. (2003). Introduction. In N. Griffin (Ed.), *The Cambridge Companion to Bertrand Russell* (pp. 1–50). Cambridge: University Press.

Husserl, E. (1900–1). *Logical investigations*. (J. N. Findlay, Trans., with Preface by M. Dummett, and edited and with introduction by D. Moran, 2001) London, New York: Routledge.

Hylton, P. (1990). *Russell, idealism and the emergence of analytic philosophy*. Oxford: Clarendon Press.

Hylton, P. (1996). Beginning with analysis. (Reprinted in *Propositions, functions, and analysis*, pp. 30–48, by P. Hylton, 2005, Oxford: Clarendon Press).

[30]Vann McGee (2004, p. 621) presents Russell's (1923) account of the vagueness of the name "Ebenezer Wilkes Smith" as a case of "the inscrutability of reference" consistent with "the line of argument of argument of Chapter Two of *Word and Object*". I have argued, in effect, that in his earlier account of the vagueness of numerical terms—specifically, in his acknowledging that if there are indefinables of the sort he previously took cardinal numbers to be, then either those indefinables, or classes of similar classes will enable us to sustain the same "formulae of arithmetic"—Russell is likewise arguing for a case of the "inscrutability of reference".

[31]An earlier version of this paper was presented at the 2013 meeting of the Society for the Study of the History of Analytic Philosophy at the University of Indiana, Bloomington. Thanks to the audience there, in particular Kevin Klement, Gregory Landini, Ian Proops, and Thomas Ricketts, for helpful comments. Thanks especially to Sébastien Gandon and Peter Hylton for reading and commenting on that earlier version.

Hylton, P. (2004). Quine on reference and ontology. In R. Gibson (Ed.), *The Cambridge Companion to Quine* (pp. 115–150). Cambridge: University Press.
Hylton, P. (2007). "On Denoting" and the idea of a logically perfect language. In M. Beaney (Ed.), *The analytic turn* (pp. 91–106). New York, London: Routledge.
Keefe, R., & Smith, P. (1997). Introduction: Theories of Vagueness. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 1–57). Cambridge, MA: The MIT Press.
Kripke, S. (2005). Russell's notion of scope. *Mind, 114*, 1005–1037.
Levine, J. (2009). From Moore to Peano to Watson: The mathematical roots of Russell's naturalism and behaviorism. *Baltic International Yearbook of Cognition, Logic and Cognition*, *4*, 1–126. http://newprairiepress.org/biyclc/vol4/iss1/10/
McGee, V. (2004). The many lives of Ebenezer Wilkes Smith. In G. Link (Ed.), *One hundred years of Russell's paradox* (pp. 611–623). Berlin, New York: Walter de Gruyter.
Moore, G. E. (1901). Identity. (Reprinted in *G.E. Moore: The early essays,* pp. 121–145, ed. by T. Regan, 1986, Philadelphia: Temple University Press).
Moore, G. E. (1903a). *Principia Ethica*. Cambridge: University Press.
Moore, G. E. (1903b). The refutation of idealism. (Reprinted in *Philosophical studies*, pp. 1–30, by G. E. Moore, 1922, London: Routledge & Kegan Paul Ltd.).
Moore, G. E. (1905). Russell's *Principles of mathematics*. Unpublished typescript in Bertrand Russell Archives.
Moore, G. E. (1942). A reply to my critics. In P. A. Schilpp (Ed.), *The philosophy of G.E. Moore* (pp. 535–677). New York: Tudor Publishing Company (2nd ed., 1952).
Moore, G. E. (1953). *Some main problems of philosophy*. London: George Allen & Unwin Ltd.
Petrilli, S. (2009). *Signifying and understanding: Reading the works of Victoria Welby and the signific movement*. Berlin: De Gruyter Mouton.
Quine, W. V. (1960). *Word and object*. Cambridge, Massachusetts: The MIT Press.
Quine, W. V. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
Quine, W. V. (1992). *The pursuit of truth* (revised edition). Cambridge, Massachusetts: Harvard University Press.
Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton, NJ: Princeton University Press.
Russell, B. (1899). The axioms of geometry. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 2*, pp. 394–415, N. Griffin and A. C. Lewis (Eds.), 1990, London: Routledge.)
Russell, B. (1899–1900). *The principles of mathematics*, Draft of 1899–1900. In G. H. Moore (Ed.), *The Collected Papers of Bertrand Russell: Volume 3* (1993, pp. 9–180). London and New York: Routledge.
Russell, B. (1900a). *A critical exposition of the philosophy of Leibniz*. London: George Allen & Unwin Ltd.
Russell, B. (1900b). Is position in time absolute or relative? In G. H. Moore (Ed.), *The Collected Papers of Bertrand Russell: Volume 3* (1993, pp. 222–233). London and New York: Routledge.
Russell, B. (1901a). On the notion of order. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 3*, pp. 287–309, G. H. Moore (Ed.), 1993, London and New York: Routledge.)
Russell, B. (1901b). The notion of order and absolute position in space and time. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 3*, pp. 241–258, G. H. Moore (Ed.), 1993, London and New York: Routledge.)
Russell, B. (1901c). Is position in time and space absolute or relative? (Reprinted in *The Collected Papers of Bertrand Russell: Volume 3*, pp. 261–282, G. H. Moore (Ed.), 1993, London and New York: Routledge.)
Russell, B. (1901d). Recent work on the principles of mathematics. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 3*, pp. 366–369, G. H. Moore (Ed.), 1993, London and New York: Routledge.)

Russell, B. (1903). *The principles of mathematics*. Cambridge: Cambridge University Press (2nd ed., 1937).
Russell, B. (1905). On denoting. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 4*, pp. 414–427, A. Urquhart (Ed.), 1994, London and New York: Routledge.)
Russell, B. (1905–06). On some difficulties in the theory of transfinite numbers and order types. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 5*, pp. 62–89, G. H. Moore (Ed.). 2014, London and New York: Routledge.)
Russell, B. (1909). Pragmatism. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 6*, pp. 260–284, J. G. Slater (Ed.), 1992, London: Routledge.)
Russell, B. (1910). The theory of logical types. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 6*, pp. 4–31, J. G. Slater (Ed.), 1992, London: Routledge.)
Russell, B. (1911a). Knowledge by acquaintance and knowledge by description. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 6*, pp. 148–161, J. G. Slater (Ed.), 1992, London: Routledge.)
Russell, B. (1911b). Analytic realism. Reprinted (A. Vellino Trans.). (Reprinted in *The Collected Papers of Bertrand Russell: Volume 6*, pp. 133–146, J. G. Slater (Ed.), 1992, London: Routledge.)
Russell, B. (1912). *The problems of philosophy*. Oxford: Home University Library.
Russell, B. (1913). Theory of knowledge. In E. Eames (Ed.), *The Collected Papers of Bertrand Russell: Volume 7*, (1984, pp. 1–178). London: George Allen & Unwin.
Russell, B. (1914a). The relation of sense-data to physics. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 8*, pp. 5–26, J. G. Slater (Ed.), 1986, London: George Allen & Unwin.)
Russell, B. (1915). *Our knowledge of the external world*. Chicago: Open Court.
Russell, B. (1918). The philosophy of logical atomism. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 8*, pp. 160–244, J. G. Slater (Ed.), 1986, London: George Allen & Unwin.)
Russell, B. (1919a). *Introduction to mathematical philosophy*. London: George Allen & Unwin.
Russell, B. (1919b). On propositions: What they are and how they mean. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 8*, pp. 276–306, J. G. Slater (Ed.), 1986, London: George Allen & Unwin.)
Russell, B. (1920). The meaning of "meaning". (Reprinted in *The Collected Papers of Bertrand Russell: Volume 9*, pp. 88–93, J. G. Slater (Ed.), 1988, London: Unwin Hyman.)
Russell, B. (1921a). *The analysis of mind*. London: George Allen & Unwin.
Russell, B. (1921b). Introduction to Wittgenstein's *Tractatus Logico–Philosophicus*. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 9*, pp. 101–112, J. G. Slater (Ed.), 1988, London: Unwin Hyman.)
Russell, B. (1923). Vagueness. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 9*, pp. 147–154, J. G. Slater (Ed.), 1988, London: Unwin Hyman.)
Russell, B. (1924). Logical atomism. (Reprinted in *The Collected Papers of Bertrand Russell: Volume 9*, pp. 162–179, J. G. Slater (Ed.), 1988, London: Unwin Hyman.)
Russell, B. (1926). [Review of *The Meaning of Meaning*] (by Ogden and Richards). (Reprinted in *The Collected Papers of Bertrand Russell: Volume 9*, pp. 138–144, J. G. Slater (Ed.), 1988, London: Unwin Hyman.)
Russell, B. (1927). *Philosophy*. New York: W.W. Norton & Company Inc.
Russell, B. (1940). *An inquiry into meaning and truth*. London: George Allen & Unwin Ltd.
Russell, B. (1944). My mental development. In P. A. Schilpp (Ed.), *The philosophy of Bertrand Russell* (pp. 3–20). La Salle, Illinois: Open Court.
Russell, B. (1948). *Human knowledge: Its scope and limits*. New York: Simon and Schuster.
Russell, B. (1959). *My philosophical development*. New York: Simon & Schuster.
Russell, B. (1968). *The autobiography of Bertrand Russell: 1914–1968*. London: George Allen & Unwin Ltd.

Schmid, A. F. (Ed.) (2001). *Bertrand Russell Correspondence sur la Philosophie, la Logicqe et la Politque avec Louis Couturat* (Vol. 1). Paris: Kimé.
Szabó, Z. G. (2005). The loss of uniqueness. *Mind, 114*, 1185–1222.
Whitehead, A. N., & Russell, B. (1910). *Principia Mathematica* (Vol. I). Cambridge: Cambridge University Press.
Whitehead, A. N., & Russell, B. (1912). *Principia Mathematica* (Vol. II). Cambridge: Cambridge University Press.
Williamson, T. (1994). *Vagueness*. London, New York: Routledge.
Wittgenstein, L. (1921). *Logisch-Philosophische Abhandlung*. *Annalen der Naturphilosophie* (References are to the English translation by D. F. Pears & B. F. McGuinness, 1961 (2nd ed., 1974), London: Routledge & Kegan Paul Ltd.).
Wittgenstein, L. (1958). *The blue and the brown books*. New York: Harper & Row.
Wittgenstein, L (1973). In G. H. von Wright (Ed.), *Letters to C.K. Ogden*. Oxford: Basil Blackwell.

Author Biography

James Levine is Associate Professor in the Philosophy Department in Trinity College Dublin. His work has focused largely on Frege, Russell, and the early Wittgenstein.

Propositional Logic from *The Principles of Mathematics* to *Principia Mathematica*

Bernard Linsky

It is sometimes said that even though Whitehead and Russell (1910), *Principia Mathematica* was one of the foundational works of Early Analytic Philosophy, at the same time it is seldom read and is even "unreadable".[1] PM is certainly not studied by contemporary philosophers beyond chapter ∗14 on definite descriptions, in part because of its antiquated notation, but also because it has been superseded by subsequent developments in logic. A contemporary logician might see the early chapters on propositional logic ∗2–∗5 as a haphazard pile of almost two hundred elementary theorems based on an unmotivated choice of axioms. Soon after the publication of the first volume of PM in 1910, Henry Sheffer and Jean Nicod had showed that propositional logic can be formulated with one connective (the "Sheffer Stroke") and then one axiom using that connective.[2] Nicod and Sheffer had been Russell's students, and their accomplishments figure centrally in the Introduction to the new second edition of *Principia Mathematica* published in 1925, giving rise to the impression that Russell repudiated the system of elementary logic in PM early on. By the mid 1920s it was known how to show that any of a number of systems of axioms for propositional logic are complete for the semantic interpretation which uses truth tables.[3] Those working in the field of Early Analytic Philosophy see in Wittgenstein's *Tractatus Logico-Philosophicus* the origins of the views that no axiomatic formulation of propositional logic is necessary and that the notion of *tautology* replaces that of *theorem* as an account of logical truth. As a result, the propositional logic of PM is thought to have been obsolete and philosophically irrelevant almost as soon as it was published. However, looking at PM in terms of its antecedents in Peano's logic and Russell's own earlier work, will help

[1] See the quotation in Griffin and Linsky (2013), at page xviii, for one statement of this view.
[2] Sheffer (1913) and Nicod (1917).
[3] Bernays (1926) includes both the result that one of the PM axioms was redundant, and the completeness of the axioms, neither of which make an appearance in the second edition of PM. See Linsky (2011).

B. Linsky (✉)
Department of Philosophy, University of Alberta, Edmonton, Canada
e-mail: blinsky@ualberta.ca

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_8

to understand many of these features that seem puzzling given what happened in logic after PM.

Russell describes hearing Giuseppe Peano and his students talking at the Paris Congress in 1900 as the "most important event" in the "most important year in my intellectual life" (Russell 1944, 12). It was this encounter with their new symbolic logic that led Russell to rewrite an almost complete draft of *Principles of Mathematics* (Russell 1993). The most important impact of Peano on Russell's project in the foundation of mathematics was in leading Russell to develop the logic of relations, and then to adopt the "Frege-Russell" definition of numbers as classes of equinumerous classes (Russell 1901). Russell needed to find definitions of mathematical notions in terms of logical notions and then prove them from logic alone. He was inspired to formulate the logic to be used in these derivations on the model of Peano's symbolic logic. Russell not only learned elements of symbolism from Peano, such as the familiar "horseshoe" ('$\supset$') for material implication and the use of dots for punctuation, and the symbolic representation of relations, but also many of the axioms and theorems that he later reworked into the system of *Principia Mathematica*. Below I approach the early chapters of *Principia Mathematica* as the result of a slow and laborious development out of Peano's original ideas. This development resulted not only in the logic of classes and relations that is crucial to the logicist project of reconstructing mathematics, and also in the slow development of the elementary logic of propositions and quantifiers which is used throughout PM. I will illustrate this development by studying a theorem that is *not* proved in those early chapters of PM.

"Peirce's Law" is a valid formula of elementary propositional logic:[4]

$$[(p \supset q) \supset p] \supset p$$

The "law" is not a theorem in *Principia Mathematica* (PM) although it is one of the axioms of *The Principles of Mathematics* (PoM) (Russell 1903) and it is proved as a theorem in Russell's paper, "Theory of Implication"(TI) (Russell 1906). Although it is not proved in *Principia Mathematica*, the two lemmas from the proof in TI are in PM, as theorems $*2 \cdot 43$ and $*2 \cdot 69$, and Peirce's Law could have been derived by the same proof as the very next theorem, $2 \cdot 7$. It thus seems likely that in Peirce's Law did in fact exist with the number $2 \cdot 7$ in an early draft of PM, but that it was removed later, as Whitehead and Russell trimmed down the theorems to eliminate what was not needed for the later development of the work.

Tracing the history of this missing theorem through earlier formulations of propositional logic helps to explain some of the features of the presentation of propositional logic in PM that seem unusual to the modern logician. Numerous aspects of the system, from the choice of primitive connectives, to the selection of axioms, and then the final choice of theorems to prove, are not explained. It is striking that the five axioms of PM are stated using the connectives '$\vee$' (' …or…')

[4]Peirce (1885) introduces the axiom. There is no evidence that Russell learned of it from Peirce.

and '$\supset$' (the material conditional: 'if…then…'), even though the conditional is a defined connective in PM. The primitive connectives in PM are instead $\sim$ ('not'), and '$\vee$', with the conditional $p \supset q$ defined as $\sim p \vee q$ in the first proposition $*1 \cdot 01$. Another peculiarity of the the system occurs with the very next proposition, the first "primitive proposition", or axiom, of this system of symbolic logic is "Anything implied by a true elementary proposition is true." This used as the rule of inference *modus ponens* in the rest of the work, despite being stated as a "Primitive Proposition" or axiom, and being expressed in words, instead of the symbols of the rest of the body of PM. What explains the particular choice of theorems that are proved in the early sections of PM? There is no hint how the proofs were discovered, or how one might try to come up with proofs of other theorems. The axioms seem to be an arbitrary list, especially when it is seen that one of them is redundant. How did Whitehead and Russell come up with these "primitive propositions"?

A study of the history of the three systems of propositional logic in PoM, TI and PM shows that there is a steady development in Russell's views about elementary logic. It appears that Russell worked out the proofs of various theorems of propositional logic that he got from Peano, and kept some of the proofs in mind when altering his systems, and even altering the choice of primitive connectives. This required the sort of attention to the details of proofs in propositional logic that in recent years has only be applied by relevance logicians when trying to find exactly what axioms are needed for particular results from different systems.[5] So, the preservation of a proof once found became a practical measure, rather than beginning again from primitives with each new choice of connectives and axioms. When composing PM, Whitehead and Russell preserved only those hard won results that were in fact used in later in proofs in combination with inferences involving quantification.[6]

An examination of the systems of PoM, TI which traces the history of Peirce's Law will help to explain these and other features of the changes in the presentation of propositional logic between *Principles of Mathematics* and *Principia Mathematica*.

The Principles of Mathematics

Section 14 on "The Propositional Calculus" begins with the assertion that this part of logic deals with the relation of implication which holds between propositions:

[5]See the introductory motivation for Relevance Logic in Anderson and Belnap (1975), Chap. 1.

[6]Andrew Tedder (Personal Communication) reports that the proofs of $*2$–$*5$ are either complete (and "gapless") or only skip steps that can be reproduced easily as the steps skipped are presented in detail in earlier proofs. The theorems are almost all either previously presented as axioms or theorems in PI, or used later in PM. (We have been so far unable to so classify only 4 of the 66 theorems of $*2$.) Each of the 27 theorems of $*3$ and 59 theorems in $*4$ belong to one of those two categories. In $*5$ only 3 of 38 theorems appear not to be in TI or used later.

> §14. The propositional logic is characterized by the fact that all its propositions have as hypothesis and as consequent the assertion of a material implication. Usually, the hypothesis is of the form "p implies p," etc., which (§16) is equivalent to the assertion that the letters which occur in the consequent are propositions. Thus the consequents consist of propositional functions which are true of all propositions.

The primitive notion of *implication* comes in two kinds, as material implication, a relation between propositions, and as formal implication, a relation between propositional functions.[7]

> §15. Our calculus studies the relation of *implication* between propositions. This relation must be distinguished from the relation of *formal* implication, which holds between propositional functions when the one implies the other for all values of the variable. Formal implication is also involved in this calculus, but it is not explicitly studied: we do not study propositional functions in general, but only certain definite propositional functions which occur in the propositions of our calculus.

At the time of writing PoM it appears that Russell only thought he needed to develop a general logic of propositions, and could deal with logical propositions concerning specific propositional functions as they were needed. Readers of PM will be familiar with the expression of formal implication involving individuals with the subscripted individual variable as in: $\phi x \supset_x \psi x$ for the expression that all ϕs are ψ. Formal implication in the logic of propositions will correspondingly involve bound variables for propositions. Consequently the first axiom of PoM will be symbolized as: $(p \supset q) \supset_{p,q} (p \supset q)$. Russell reads this as: "If p implies q, then p implies q, whatever p and q may be." It expresses a "formal implication" involving propositions.

Russell does not make a distinction between quantification over propositions and over individuals, as both are in the range of all quantifiers. Conditionals restricted to certain entities required an antecedent, as is familiar from saying that "All ϕs are ψs" is represented as "Everything is such that if it is ϕ then it is ψ" Even an assertion which involves quantifying over propositions rather than individuals will be restricted to propositions with such an antecedent hypothesis.[8]

Section 16 contains this:

> §16 … It may be observed that, although implication is indefinable, *proposition* can be defined. Every proposition implies itself, and whatever is not a proposition implies nothing. Hence to say "p is a proposition" is equivalent to saying "p implies p"; and this equivalence may be used to define propositions.

Using the notion of material implication Russell is able to define conjunction, which he calls "the logical product of two propositions". The product of p and q is written by writing one after the other as pq, although to mark the scope of conjunctions it is more convenient to use the dot, $p \cdot q$ as later in PM:

---

[7]That it is material implication still does not make implication a truth-functional connective, giving a truth value as a function of the truth-values of its constituents. Instead the value of '$p \supset q$' is a *proposition* which differs as p and q differ.

[8]Gregory Landini speaks of Russell as having only one "style of variable". Landini(1996) and elsewhere.

> §18 … If p implies p, then, if q implies q, pq, (the logical product of p and q) means that if p implies that q implies r then r is true.

As a result the axioms can all be stated using only the symbol '$\supset$' for material implication, with the formula '$(p \supset p)$' expressing the hypothesis that p is a proposition. However Russell also uses the notion of "logical product" in the statement of his axioms. This is no different in principle from what is done in *Principia* except there it is not only axioms, but even the rule of inference that are stated for the defined notion, which is material implication, or '$\supset$'. This does look quite odd, but we see that it is just not part of Russell's conception of axioms for logic, from the very beginning.

We are now able to express the ten primitive propositions of *Principles of Mathematics* from §18 in terms of the primitive connective and the defined connective for "and". Russell says "… of these, all except the last will be found in Peano's accounts of the subject."[9]

1. $(p \supset q) \supset (p \supset q)$
2. $(p \supset q) \supset (p \supset p)$
3. $(p \supset q) \supset (q \supset q)$
4. A true hypothesis in an implication may be dropped and the consequent asserted.
5. $(p \supset p) \cdot (q \supset q) \supset (pq \supset p)$ *Simplification*
6. $(p \supset q) \cdot (q \supset r) \supset (p \supset r)$ *Syllogism*
7. $\{(q \supset q) \cdot (r \supset r) \cdot [p \supset (q \supset r)]\} \supset (pq \supset r)$ *Importation*
8. $[(p \supset p) \cdot (q \supset q)] \supset \{(pq \supset r) \supset [p \supset (q \supset r)]\}$ *Exportation*
9. $(p \supset q) \cdot (q \supset r) \supset (p \supset qr)$ *Composition*
10. $(p \supset p) \cdot (q \supset q) \supset \{[(p \supset q) \supset p] \supset p\}$ *Reduction*

Thus already two of the oddities of *Principia Mathematica* can be better understood by seeing its ancestor in the system in *The Principles of Mathematics*. The statement of the primitive propositions in PM in terms of a defined connective ('.') is one. If all the propositions of propositional logic state implications where both the antecedent and consequent are implications then this means that logic is the study of implications, with other connectives to be understood in terms of implication.[10] While Russell changed his mind about the primitive notions of logic in PI and then in PM, the origins of his views in PoM help to explain why he kept the principles expressed in terms of $\supset$ as long as possible in the primitive propositions.

That implication is the only primitive notion of logic also helps to understand another peculiarity of PM, namely Russell's reluctance to express Primitive Proposition $*1 \cdot 1$ symbolically, but instead with an English sentence. He says of the antecedent rule in PoM, Proposition 4, that:

---

[9]The last is expressed this way: If p implies p and q implies q, then "'p implies q', implies p" implies p. The names also come from Peano, with the exception of "Reduction".

[10]See the discussion in Byrd (1989).

> §18 This is a principle incapable of formal symbolic statement, and illustrating the essential limitations of a formalism — a point to which I will return at a later stage.

The "limitation of a formalism" that concerns Russell is complicated. In part it is that his symbolism does not allow him to express the difference between a rule of inference and a proposition in the system. The "later stage" at which Russell returns to the problem includes this:

> §38 The independence of this principle is brought out by a consideration of Lewis Caroll's puzzle, "What the Tortoise said to Achilles." The principles of inference which we accepted lead to a proposition that, if p and q be propositions, then p together with "p implies q" implies q. At first sight, it might be thought that this would enable us to assert q provided p is true and implies q. But the puzzle is question shows that this is not the case, and that until we have some principle, we shall only be led to an endless regress of more and more complicated implications, without ever arriving at the assertion of q. We need the notion of *therefore*, which is quite different from the notion of *implies*, and holds between different entities.

Russell takes the point of "What the Tortoise said to Achilles" (Caroll 1895) to be that a system of logic cannot consist solely of axioms, for stating the rules of inference as axioms still requires a rule for the application of those axioms, leading to an infinite regress. The difference between a rule of inference and a theorem, however, does not seem to be beyond the power of symbolic expression. Using Frege's "assertion sign" ('$\vdash$'), which Russell himself mentions, and the formulation of inference rules inherited from Gentzen we can easily express a rule of inference as follows:

$$\frac{\vdash p \quad \vdash p \supset q}{\vdash q}$$

This, however, is a rule of inference stating what follows from previously proved statements. It says that if p is a theorem, and $p \supset q$ is a theorem, then one can prove q as the next step in a derivation. It does not say anything about what can be proved from an assumption. The principal objection to the possibility of expressing the primitive proposition symbolically is expressed just before the reference to Lewis Caroll:

> §38. One of our indemonstrable principles was, it will be remembered, that if the hypothesis in an implication is true, it may be dropped, and the consequent asserted. This principle, it was observed, eludes formal statement and points to a certain failure of formalism in general. The principle is employed whenever a proposition is said to be *proved*; for what happens is, in all such cases, that the proposition is shown to be implied by some true proposition. Another form in which the principle is constantly employed is the substitution of a constant, satisfying the hypothesis, is the consequent of a formal implication. If ϕx implies ψx for all values of x, and if a is a constant satisfying ϕx, we can assert ψa, dropping the true hypothesis ϕa. This occurs, for example, whenever any of these rules of inference which employ the hypothesis that the variables involved are propositions, are applied to particular propositions.

Here Russell points out a second aspect of the fourth primitive proposition which also suggest that it is "incapable of formal symbolic statement", that it formulates a rule of inference which applies to "hypotheses", that is to a proposition that is not

proved as a theorem. In a natural deduction formulation of logic we may assume a proposition for the purposes of a further argument, such as showing that the assumption is false by *reductio ad absurdum*. Again the symbolism of proof theory can be extended to express this further relation of *implication* as a *consequence* relation, between as set of assumptions Γ and a consequence of those assumptions p symbolised as: '$\Gamma \vdash p$'. The primitive proposition then becomes:

$$\frac{\Gamma \vdash p \quad \Gamma \vdash p \supset q}{\Gamma \vdash q}$$

This expresses the notion of "proving" a proposition by showing that it is implied by a "true" proposition. We must distinguish between a "true proposition" and a "theorem" in the statement of this primitive proposition, and in *Principia Mathematica*. Still, however, the failure of "symbolism" is simply a failure of Russell's symbolism, as two further notions, developed later in the history of logic, do enable us to express Russell's primitive propositions in symbols.

A third role that the primitive proposition plays that Russell believes incapable of symbolization is its role as a rule of *universal instantiation*: "the substitution of a constant, satisfying the hypothesis, is the consequent of a formal implication." That is: "If ϕx implies ψx for all values of x, and if a is a constant satisfying ϕx, we can assert ψa." Once more, this is a fundamental rule of contemporary symbolic logic, which can be expressed as:

$$\frac{\Gamma \vdash \phi a \quad \Gamma \vdash (x)\phi x \supset \psi x}{\Gamma \vdash \psi a}$$

We see then that Primitive Proposition 4 is also the quantifier rule for the system of *Principles of Mathematics*, both for propositional logic and the unformulated logic of quantifiers. It is possible to express each of these three roles for Proposition 4 in a formalism, however that requires three developments of symbolism that Russell did not appreciate even through the writing of PM. He was simply wrong to claim that they are "incapable of formalism".

The tenth primitive proposition in PoM is well known as "Peirce's Law", as it was first proposed as an axiom of propositional logic by Peirce in (1885). Russell makes no reference to Peirce, however, but instead describes this last axiom as follows:

> §18 … … it has less self evidence than the previous principles, but is equivalent to many propositions that are self-evident. … The principle is especially useful in conjunction with negation. Without its help, by means of the first nine principles, we can prove the law of contradiction; we can prove, if p and q be propositions, that p implies not-not-p… But we cannot prove without reduction or some equivalent (so far at least as I have been able to discover) that p or not-p must be true (the law of excluded middle); that ever proposition is equivalent to the negation of some other proposition; that not-not-p implies p, or that "p implies q" implies "q or not- p". Each of these assumptions is equivalent to the principle of reduction, and may, if we choose, be substituted for it.

So Russell adds Reduction as a primitive proposition firstly because it is the necessary in order to prove various propositions concerning negation, and secondly because it is expressed in terms of implication, as are the other primitive propositions. Russell understands the material conditional $p \supset q$ to be equivalent to $q \vee \sim p$ but neither $\vee$ nor $\sim$ is a primitive connective of PoM. In PM, where those are the primitive connectives, and $\supset$ is defined in terms of them, he still maintains the preference for stating as many primitive propositions as possible in terms of $\supset$.

Disjunction is defined at §19 in terms of the material conditional:

> Disjunction or logical addition is defined as follows: "p or q" is equivalent to "'p implies q' implies q".

This is readily formalized using a formula equivalent to a disjunction:

$$p \vee q = (p \supset q) \supset q \quad \text{Df.}$$

It is striking from the appearance of the primitive propositions in PoM that there are no negations. Negation, however, is defined in §19:

> §19 … Hence we may proceed to the definition of negation: not-p is equivalent to the assertion that p implies all propositions, i.e. that "r implies r" implies "p implies r" whatever r may be.

At the time Russell composed PoM, that is before he encountered Frege's notion of quantification in June of 1902, all quantification in logic took the form of *formal implication* or as universally quantified conditionals. Recall that furthermore these quantifiers ranged over both individuals and propositions. This is called the use of only "one style of variable" in Landini's (1996) discussions of Russell's logic. As a result this definition of negation would be symbolized as:

$$\sim p = p \supset_r r \quad \text{Df.}$$

Because formal implication is the means by which propositional quantification is expressed, although it is more striking to say that the propositional logic of PoM has one primitive notion of implication, as Russell does, it is more informative to contemporary logicians to say that there are *two* primitive notions, material implication ($\supset$) and propositional quantification.

Conjunction, which is used in the statement of the axioms, is also defined using formal implication:

$$p \cdot q = \{[p \supset (q \supset r)] \supset_r r\} \quad \text{Df.}$$

Russell asserts in §18 that without Reduction he is unable to prove various fundamental principles about negation, such as the law of excluded middle, or double negation ($\sim \sim p \supset p$), and others. The law of the excluded middle, $p \vee \sim p$ (LEM) in the primitive notation of *Principles* would be, replacing the symbol $\vee$ by its definition as:

$(p \supset \sim p) \supset \sim p$, then replacing $\sim p$ with its definition, becomes:

$$(p \supset (p \supset_r r)) \supset (p \supset_r r)$$

How the law of excluded middle might be proved from Reduction can be seen most easily in an extended language which includes the defined expressions $\sim$, $\vee$ and a further propositional constant, $\perp$, the "*falsum*". In such a language the following is an instance of Reduction:

$$\{[(p \vee \sim p) \supset \perp] \supset (p \vee \sim p)\} \supset (p \vee \sim p)$$

The antecedent can be proved since $p \supset \perp$ follows from $(p \vee \sim p) \supset \perp$. $p \supset \perp$ is by definition $\sim p$, and so we can prove that $(p \supset \perp) \supset (p \vee \sim p)$. This is the antecedent of the instance of Reduction stated above and LEM the consequent. LEM follows from this by *modus ponens*.[11]

Proving this in the language and system of *Principles* would require expressing the law of the excluded middle without $\sim$, $\vee$ or $\perp$. The *falsum* constant could be replaced by a specific contradiction, say $p \cdot \sim p$ which will have the same logical power as $\perp$, in particular, that $p \supset (p \cdot \sim p)$ will amount to $\sim p$. First, replacing $\perp$, the instance of Reduction with which we begin is still recognizable as such:

$$\{[(p \vee \sim p) \supset (p \cdot \sim p)] \supset (p \vee \sim p)\} \supset (p \vee \sim p)$$

Next, replacing each instance of $\sim p$ by it's definiendum $(p \supset_r r)$, we get:

$$\{[(p \vee (p \supset_r r)) \supset (p \cdot (p \supset_r r))] \supset (p \vee (p \supset_r r))\} \supset (p \vee (p \supset_r r))$$

The final formula, after replacing the $\vee$ with its definition, is still an instance of Reduction, even if it is unreadable:

$$\begin{aligned}\{[([\,p \supset (p \supset_r r)] \supset (p \supset_r r)) \supset (p \cdot (p \supset_r r))] \\ \supset ([p \supset (p \supset_r r)] \supset (p \supset_r r))\} \\ \supset ([p \supset (p \supset_r r)] \supset (p \supset_r r))\end{aligned}$$

It is clear that Russell must have conducted his derivations in a language with defined symbols and not in the language with only implication and conjunction.

Double negation probably can likely be proved from another instance of Reduction:

$$\{[(\sim \sim p \supset p) \supset \perp] \supset ((\sim \sim p \supset p)\} \supset ((\sim \sim p \supset p))$$

[11]Landini (1996) provides PoM with a system that includes the "falsum" '$\perp$' (defined as '$p \supset p.\ \supset_p p$') and other technical additions. He shows that the expanded system is complete.

In this case the antecedent seems to follow directly, for $(\sim\sim p \supset p) \supset \perp$ is by definition $\sim(\sim\sim p \supset p)$, which is equivalent to $\sim\sim p \cdot \sim p$, and from that contradiction it should be possible to prove anything, including $(\sim\sim p \supset p)$ and so the antecedent of the instance of Reduction of which $(\sim\sim p \supset p)$ is the conclusion.

Yet the question remains, how did Russell discover the principle of "Reduction" on his own, if in fact he did not discover it from reading Peirce? It might be that he saw that the procedure used above is generally a good way of proving theorems of propositional logic. Take some principle L, then it is likely easy to prove that $(L \supset \perp) \supset L$. Then, by Reduction and *modus ponens*, we can deduce L.

Another hypothesis is that Russell stumbled onto Reduction while trying to prove LEM which will be expressed as: $(p \supset (p \supset_r r)) \supset (p \supset_r r)$. The law of excluded middle seems to be a direct result of universal quantification with respect to the propositional variable r, distributed over the conditional, from the propositional principle known as "Contraction":

$$(p \supset (p \supset r)) \supset (p \supset r)$$

It is not obvious, however, how Reduction is needed for a proof of Contraction.

In the language with $\sim$, LEM is defined as

$$(p \supset \sim p) \supset \sim p$$

This is logically equivalent to $\sim p \vee p$, or, by definition:

$$(\sim p \supset p) \supset p$$

This principle is not valid intuitionistically or in other systems with weak principles concerning negation, as it would seem to validate an indirect proof of p from a certain inference based on assuming that p is absurd. When expressed with the definition of PoM this is:

$$[(p \supset_r r) \supset p] \supset p$$

Reduction might mistakenly look like an instance of this (logically equivalent) principle.[12] While Reduction may be needed as a primitive proposition to prove these results, because all the primitive propositions are expressed with material and formal implication for propositions, it is still a surprise that it is the only axiom needed to make the system adequate for proving the principles of propositional logic involving negation. However, once it is seen that LEM and double negation

[12]It would be an instance of this alternative, with a different scope for the propositional quantifier: $[(p \supset r) \supset p] \supset_r p$. Alasdair Urquhart (personal communication) has speculated about this as the source of Russell's rediscovery of Peirce's Law. Gregory Landini (personal communication) has pointed out that these are in fact logically equivalent despite the different scopes of the quantification of r.

can be proved as above, it is obvious that the ten primitive propositions of PoM form a complete set of axioms for propositional logic.

The Theory of Implication

In 1906 Russell published a different formulation of propositional logic, embedded in a theory of quantificational logic. This theory was couched in terms of the "substitutional theory" that Russell was developing at that point in which propositions and individuals were still equally within the range of the quantifiers and so now prefixed to formulas in Frege's style, rather than subscripted as in formal implication. In addition to explicit quantification, the other primitives were material implication $\supset$ and negation $\sim$. He says of his proposal to take negation as a primitive (Russell 1906, 60):

> Instead of taking negation as a primitive idea, as was done in $*1$, it is possible to regard the property stated in $*7 \cdot 22$ as giving the *definition* of negation, *i.e.*, we may put:
>
> $$\sim p. =: p. \supset .(s).s \quad \text{Df}$$

Adopting negation as a primitive notion requires the addition of three axioms explicitly concerning negations, although now Reduction can be proved as a theorem. The cost of adding a new primitive "idea" as well as three new axioms to replace one before is justified as follows:

> My reason for not adopting this method is not its artificiality or its difficulty, but the fact that it never enables us to know that anything whatever is *false*. It enables us to prove the *truth* of whatever can be proved true by the method adopted above, and it does not enable us to prove the truth of anything which is in fact false. It even enables us to prove, concerning all the propositions which can be proved false by the above method, that, if they are true, then everything is true; but if any man is so credulous as to believe that everything is true, then the method in question is powerless to refute him. For example, we get the law on contradiction in the form
>
> $$\vdash : p. \sim p \ . \supset \ . \ (s)s;$$
>
> but this does not show that $p. \sim p$ is false, unless we assume that $(s). \ s$ is false. (Russell 1906, 60)

Russell is able to adopt this definition of '$\sim p$' as '$p \supset (s). s$' since he had by this time come to adopt Frege's analysis of quantifiers as applying to formulas rather than hidden in the expression of formal implication, as in '$p \supset_s s$'.

In TI Russell is thus able to define the propositional constant '$\bot$' with '$(s). s$', however finds the definition of negation '$\sim p$' in terms of that constant '$p \supset \bot$' to be inadequate, for the characteristic property of $\bot$, that it implies every proposition, is not enough to capture negation.

Russell explicitly recognizes the use of a rule of substitution in TI, by which instances of theorems are immediate consequences. Implication can still be read as

a relation, and so the restriction on primitive propositions to propositions as antecedents and consequents can be avoided by simply holding that a material implication is only true if the consequent is a true proposition or the antecedent not a true proposition. It is not necessary to restrict assertions to propositions with the antecedent '$p \supset p$' as in PoM, even though implication is still seen as a relation between propositions.[13]

The axioms for propositional logic, though now much more familiar in appearance, are still ten in number:[14]

[TI] $*2 \cdot 1$ Anything implied by a true proposition is true.
[TI] $*2 \cdot 2$ If a propositional function $(C \{\!|\, y)$ is true for *any* value of y, it is true for such and such a value.
[TI] $*2 \cdot 5 \vdash p \supset p \quad ID$
[TI] $*2 \cdot 6 \vdash p \,.\supset. q \supset p \quad Simp$
[TI] $*2 \cdot 7 \vdash :. p \supset q .\supset: q \supset r .\supset. p \supset r \; Syll$
[TI] $*2 \cdot 8 \vdash p .\supset. q \supset r :\supset: q .\supset. p \supset r \; Comm$
[TI] $*2 \cdot 9 \vdash .\sim(\sim p) \supset p \quad Neg$
[TI] $*2 \cdot 91 \vdash :p \supset \sim p .\supset. \sim p \quad Abs$
[TI] $*2 \cdot 92 \vdash :p \supset \sim q .\supset. q \supset \sim p \quad Transp$

As before, the English sentence [TI] $*2 \cdot 1$ states the rule of *modus ponens*. The second primitive proposition, [TI]$*2 \cdot 2$, serves to allow a rule of universal instantiation, although the quantifiers range over both propositions and individuals. The three quantifier rules explicitly formulated for propositional functions which apply to individuals, including instantiation, generalization and a rule for distributing quantifiers over the conditional are stated later. As TI was written during Russell's attachment to the so-called "substitutional" logic, which has all quantifiers ranging over propositions (as well as individuals), this first quantificational rule should be read as applying to propositions only. As a rule replacing a bound (apparent) variable ranging over propositions by an instance, it amounts to a rule of substitution. If some formula $\ldots p \ldots$ is provable with p a bound variable, then an instance of it, say for $q \supset r$, will be a result of instantiation of $q \supset r$ for the bound variable p. In PM, Whitehead and Russell try to avoid the use of bound propositional variables, and so are not able to state a principle of substitution in this form. Russell later expressed regret that no alternative rule was proposed, and so we have yet another of the oddities of PM as a system of logic: it contains no explicit rule of substitution, despite the frequent use of such a rule from the very beginning of $*2$.[15]

With negation as a primitive connective in addition to the material conditional, there is no need for Reduction in order to prove the properties of negation as in

[13]Allen Hazen, in conversation, has put this by saying that individuals are treated as "honorary (false) propositions."

[14]Numbered formulas will be prefixed with '[TI]' to indicate their origin in "Theory of Implication".

[15]See the note on page 151 of *Introduction to Mathematical Philosophy* (Russell 1919).

PoM. Using the principles *Neg*, *Abs* and *Trans* as primitive propositions the properties that Russell mentions in PoM are proved as theorems in TI, including Reduction:

$$[\text{TI}] * 3 \cdot 51 \vdash : . p \supset q . \supset . p : \supset . p$$

The proof of Reduction is a direct application of two earlier proved theorems as lemmas. The first is:

$$[\text{TI}] * 3 \cdot 47 \vdash : . p \supset q . \supset . q : \supset : q \supset p . \supset . p$$

Notice that by the definition of '$\vee$' used in PoM, this is equivalent to $p \vee q . \supset . q \vee p$. Asserting that $\vee$, so defined, is commutative, this is fundamental principle of logic independently of its further use. An instance of [TI] $*3 \cdot 47$ obtained by replacing 'q' by '$p \supset q$' is the next step in the proof:

$$p \supset (p \supset q) . \supset . (p \supset q) : \supset : . p \supset q . \supset . p : \supset . p$$

The consequent of this conditional is Reduction. Its antecedent is the second lemma:

$$[\text{TI}] * 3 \cdot 5 \vdash : . p . \supset . p \supset q : \supset p \supset q$$

While this lemma is of considerable inference for the study of relevant conditionals (where it is called "Contraction"), it is used only in this proof. One step of *modus ponens* results in a proof of Reduction.

It might appear from this that Russell was wrong to think that Reduction was needed in order to prove theorems about negation, for neither [TI] $*3 \cdot 47$ nor [TI] $*3 \cdot 5$ involves '$\sim$', and could have been used as primitive propositions. Tracing through the proof of both of these lemmas leads, through [TI] $*3 \cdot 42$ to:

$$[\text{TI}] * 3 \cdot 41 \vdash : \sim p \supset p . \supset . p$$

After the proof of $*3 \cdot 41$ Russell remarks:

> The above is the first use of $*2 \cdot 91$. It asserts that a proposition must be true if it can be deduced from the supposition that it is false.

Russell thus signals that this theorem requires primitive proposition [TI] $*2 \cdot 91$ ("*Abs*"), which makes use of negation. He thus makes it abundantly clear that the later proof of Reduction relies on the properties of negation, confirming his remarks about it in PoM.[16]

[16]This point is taken from Allen Hazen in Restall (2000, 43).

After its proof as [TI] $*3 \cdot 51$, Reduction is not used in any further proofs, although its preceding lemma, [TI] $*3 \cdot 5$, the principle of contraction, is used repeatedly in what follows.

Principia Mathematica

The primitive connectives of PM are '$\vee$' and '$\sim$' with the material conditional '$p \supset q$' now defined as '$\sim p \vee q$' in the first proposition:

$$*1 \cdot 01.\ p \supset q\,.=.\sim p \vee q \qquad \text{Df.}$$

The new list of primitive propositions is shorter than in TI. There are two primitive propositions which express rules of inference and are not symbolized. The first is:

$*1 \cdot 1$. Anything implied by a true elementary proposition is true. Pp.

Of this Whitehead and Russell say:

> We cannot express the principle symbolically, partly because any symbolism in which p is variable only gives the *hypothesis* that p is true, not the *fact* that it is true. (PM, 94)

Presumably the other reason is still the problem of Lewis Caroll about expressing a rule as a proposition. But there is an additional primitive proposition, which is also not expressed in symbols:

> $*1 \cdot 11$. When ϕx can be asserted, where x is a real variable, and $\phi x \supset \psi x$ can be asserted, where x is a real variable, then ψx can be asserted, where x is a real variable. Pp.

These two propositions, between them, perform the three tasks of the unsymbolized primitive proposition 4 of PoM.

PM has no constants but only "real variables" to which we can instantiate, so this is the only formulation of *modus ponens* for quantificational logic that is possible. In PM propositions $*9 \cdot 1.\vdash :. \phi x\,.\supset .(\exists z).\,\phi z$ Pp and $*10 \cdot 1.\vdash :.(x).\,\phi x\,.\supset .\,\phi y$, are the instantiation rules for the two formulations of quantificational logic in $*9$ and $*10$.

$*1 \cdot 11$ is thus a version of *modus ponens* for use in quantificational logic, and so performs only two of the roles that Russell had thought would make it "incapable of formal symbolic statement." It still serves as an instantiation rule for *propositional* variables, which are generally avoided in PM. For example, in the discussion immediately following $*14 \cdot 3$ which makes use of p and q as *real*, i.e. bound propositional variables, Whitehead and Russell say:

> The following propositions are immediate applications of the above. They are, however, independently proved, because $*14 \cdot 3$ introduces propositions (p, q namely) as apparent variables, which we have not done elsewhere, and cannot do legitimately without the

> explicit introduction of the hierarchy of propositions with a reducibility-axiom such as $*12 \cdot 1$. (PM, 185)[17]

Primitive proposition $*1 \cdot 11$ is used in the proof of $*2 \cdot 06$, seemingly simply to justify use of *modus ponens* in the consequent of a conditional, and thus only on the basis of an "hypothesis" rather than a "known truth". However, $*1 \cdot 11$ explicitly mentions "real variables" yet $*2 \cdot 06$ only uses the variables for propositions p and q. This suggests that here we have yet another argument that there are bound and free *propositional* variables in PM, and that the propositional letters "p", "q", etc. should indeed be treated as free variables rather than as schematic letters.

The other primitive propositions are expressed with just $\supset$ and $\vee$, thus using the defined connective $\supset$ and ignoring a primitive connective $\sim$.

$$*1 \cdot 2.\ \vdash :p \vee p . \supset . p \quad \text{Pp} \quad (\text{Taut})$$
$$*1 \cdot 3.\ \vdash :q . \supset . p \vee q \quad \text{Pp} \quad (\text{Add})$$
$$*1 \cdot 4.\ \vdash :p \vee q . \supset . q \vee p \quad \text{Pp} \quad (\text{Perm})$$
$$*1 \cdot 5.\ \vdash : p \vee (q \vee r) . \supset . q \vee (p \vee r) \quad \text{Pp} \quad (\text{Assoc})$$
$$*1 \cdot 6.\ \vdash : . q \supset r . \supset : p \vee q . \supset . p \vee r \quad \text{Pp} \quad (\text{Sum})$$

(Paul Bernays (1926), in results first presented in his Habilitationsschrift in 1918, shows that this could be reduced by one, as $*1 \cdot 5$ *Assoc* can be proved from the others.[18])

That this is system is intended as an improvement on that of TI is clear from the first theorems of PM, which are in fact the primitive propositions of TI. (*Abs, Simp, Transp, Comm* and *Syll*) and the other two are proved soon after (*ID* at $*2 \cdot 08$,and *Neg* at $*2 \cdot 14$). Whitehead and Russell thus demonstrate that the same theorems can be proved with this shorter list of primitive propositions. Still, however, some of the same proofs to be carried over to PM, using lemmas where before there were primitive propositions.

The most striking case of this situation shows up with the Reduction where the lemmas are proved, without the theorem. The two lemmas from TI show up, again as theorems: [TI] $*3 \cdot 47$ is now $*2 \cdot 43$, and [TI] $*3 \cdot 5$ is now $*2 \cdot 69$. Exactly the same two step proof as in TI would result as a proof of Reduction as the next number $*2 \cdot 7$.

But there is no $*2 \cdot 7$ in PM. The next proposition after $*2 \cdot 69$ is instead the start of another chain of lemmas with $*2 \cdot 73$. It seems most likely that Whitehead and Russell had proved Reduction, along with the other primitive propositions of PoM,

[17]As the bound ("apparent") variables range over only a single type, and are not "typically ambiguous" as free ("real") variables, the use of bound variables in stating this principle would require something like the use of predicative functions for the definition of extensional contexts as they were for the definition of classes in the "no-classes" theory of $*12$.

[18]Bernays wrote to Russell about this result, although there is no mention of it in the second edition of PM. That this is not mentioned is a peculiarity of the *second edition* of PM, and so falls outside of the scope of this essay. See, however, Linsky (2011, §3.2.2).

as Russell had in TI, but found that it was again not used in what followed. The theorem was then culled from the manuscript in order to reduce the printed theorems to those that were needed for what followed. On the other hand, the two lemmas for it, as statements of "Contraction" and that the disjunction, defined with material implication, is commutative, were theorems from TI of sufficient independent interest to keep. It appears that Whitehead and Russell tried to keep as many of the proofs from TI as possible, and so wanted to prove the primitive propositions and many of the theorems of TI at an early stage in their work on PM, and only found which could be deleted later on.[19] A careful examination of the proofs in PM in comparison with those in TI will likely reveal more about the development of PM, showing that the proofs of TI were saved, when possible, and how new proofs from had otherwise to be designed. Reduction was not used later in TI either, however it had been an axiom of PoM, and was saved for that reason, as were the proofs of the axioms of TI were presented in PM even if they were not used later there.

There may be other results coming out of a careful study of the propositional logic in PM. Few logicians have looked to the early chapters of PM to find proof strategies for axiomatic formulations of propositional logic.[20]

A search for how Whitehead and Russell found the proofs in PM may result in the discovery of general techniques for finding proofs for axiomatically formulated classical propositional logic.

The early chapters in PM should be seen as the result of a cumulative effort, which had begun with Peano's work, to find primitive notions and propositions from which all theorems needed for work on the foundations of mathematics could be proved. As primitives changed, it was possible to devise new proofs, but clearly, it was always desired to keep as many of the results of previous labors as possible. Some theorems and lemmas were dropped as the subject developed. The propositional logic of PM was the result of an evolution rather than a system cleverly built anew from a unique choice of primitive notions, rules, and axioms.[21]

[19]This account is supported by the notation for deletion next to "$*2 \cdot 7$", "$*2 \cdot 71$" and "$*2 \cdot 72$" which are not marked as used later, from a list of "Propositions, where used` written in Russell's handwriting, and now in the Bertrand Russell Archives, (RA 230.031270). Only a handful of such numbers were deleted at the beginning of the list. The interest in where propositions were used was probably aimed at ensuring that all lemmas for later theorems were in fact proved earlier. Perhaps it was clear that not much space could be saved by deleting theorems not used later.

[20]For example, Hughes and Londey (1965) present proof strategies for students studying the PM system in their first symbolic logic course. Newell, Shaw and Simon (1963) used the theorems of $*2$ as a test for their automated theorem proving "Logic Theory Machine," but used only the elementary heuristics of "chaining" rather than strategies that logicians might use.

[21]My thanks to members of the Logic Reading Group at the University of Alberta, for help with all aspects of this project, Allen Hazen, Jeff Pelletier, Andrew Tedder and Hassan Masoud. Nicholas Griffin directed me to the "list of props. where used" document in the Bertrand Russell Archives.

References

Anderson, A. R., & Belnap, N. (1975). *Entailment: The logic of relevance and necessity* (Vol. I). Princeton: Princeton University Press.

Bernays, P. (1926). Axiomatische Untersuchung des Aussagen-Kalkuls der Principia Mathematica. *Mathematische Zeitschrift, 25*, 305–320.

Byrd, M. (1989). Russell Logicism, and the Choice of Logical Constants. *Notre Dame Journal of Formal Logic, 30*(3), 343–361.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind, 4*, 278–280.

Griffin, N., & Linsky, B. (2013). *The Palgrave Centenary Companion to Principia Mathematica*. Houndmills, Basingstoke: Palgrave Macmillan.

Hughes, G. E., & Londey, D. G. (1965). *The elements of formal logic*. London: Methuen.

Landini, Gregory. (1996). Logic in Russell's *principles of mathematics*. *Notre Dame Journal of Formal Logic, 37*(4), 554–584.

Linsky, Bernard. (2011). *The Evolution of Principia Mathematica*. Cambridge: Cambridge University Press.

Newell, A., Shaw J. C., & Simon N. (1963). Empirical explorations with the logic theory machine: A case study in heuristics. In E. A. Feigenbaum & J. Feldman, (Eds.), *Computers and thought* (pp. 109–133). New York: McGraw-Hill.

Nicod, Jean. (1917). A reduction in the number of the primitive propositions of logic. *Proceedings of the Cambridge Philosophical Society, 19*, 32–41.

Peirce, C. S. (1885). On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics, 7*, 180–196.

Restall, G. (2000). *An introduction to substructural logics*. London: Routledge.

Russell, B. A. (1901). The logic of relations with some applications to the theory of series. In G. H. Moore (Ed.), *The collected papers of Bertrand Russell* (Vol. 3, pp. 310–349). London: Routledge.

Russell, B. A. (1903). *The principles of mathematics.* Cambridge: University Press.

Russell, B. A. (1906). The theory of implication. *American Journal of Mathematics*, *28*, 159–202. Reprinted in *Towards Principia Mathematica*, *The Collected Papers of Bertrand Russell*, Vol. 5, pp. 14–61 by Gregory H. Moore, Ed., London, New York: Routledge, 2014.

Russell, B. A. (1919). *Introduction to mathematical philosophy*. London: George Allen & Unwin.

Russell, B. A. (1944). My mental development. In P. A. Schillp (Ed.), *The philosophy of Bertrand Russell* (pp. 3–20). LaSalle: Open Court.

Russell, B. A. (1993). The principles of mathematics, Draft of 1899-1900. In G. H. Moore (Ed.), *The collected papers of Bertrand Russell* (Vol. 3, pp. 3–180). London: Routledge.

Sheffer, H. M. (1913). A set of five independent postulates for boolean algebras, with applications to logical constants. *Transactions of the American Mathematical Society, 14*, 481–488.

Whitehead, A. N., & B.A. Russell. (1910). *Principia mathematica* (2nd ed., pp. 1925–1927). Cambridge: Cambridge University Press. (Page references to second edition.).

Author Biography

Bernard Linsky is Professor of Philosophy at the University of Alberta and a fellow of the Royal Society of Canada. He is the author of *Russell's Metaphysical Logic* and *The Evolution of Principia Mathematica* and has written articles on Russell's notion of logical construction and Russell's interactions with Frege, Meinong and Chwistek.

Part III
Wittgenstein

Later Wittgenstein on the Logicist Definition of Number

Sorin Bangu

1.

In 1939, Wittgenstein gave a series of lectures on the philosophy of mathematics at Cambridge. Alan Turing came regularly; some of the auditors took detailed notes during the thirty-one lectures, and the material was later on edited by Cora Diamond and published (originally by Cornell University Press) in 1975 under the complete title *Wittgenstein's Lectures on the Foundations of Mathematics. Cambridge 1939—From the notes of R.G. Bosanquet, Norman Malcolm, Rush Rhees, and Yorick Smythies* (LFM hereafter).

As one may imagine, "it was quite a course"—to cite from John Canfield's review of the collection (Canfield 1981, p. 333). Wittgenstein talked freely and allowed questions and interventions from the 'students', fact which, as the readers of the book will immediately appreciate, contributed significantly to clarifying his pronouncements. Less aphoristic and polyphonic than in his writings, the Wittgenstein of LFM comes unexpectedly close to the normal, conventional, discursive style of philosophizing. Yet this is not to say that his profound originality didn't shine through—on the contrary: many passages still seem impenetrable and require thorough analysis.

In the middle of lecture XVI he sets for himself the following task:

> I want to get on to a terribly difficult business – a real morass – Russell's definition of number. It seems as though, if number is defined in this way (or Frege's way), everything will be clear. (LFM, p. 156)

As the last sentence intimates, Wittgenstein is not going to be particularly happy with this definition. My aim in this paper is to understand why. Or, more modestly, to make some progress toward this goal, since, as we'll see, the discussion around

S. Bangu (✉)
University of Bergen, Bergen, Norway
e-mail: sorin.bangu@uib.no

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_9

this definition branches off in several directions, and for reasons of space I won't be able to follow them.[1] What can be done here, however, is to follow one strand of his thoughts advanced in lecture XVI: his remarks on the main conceptual ingredient of the definition, the notion of one-to-one correlation.[2] He develops his views on the issue quite substantially within this lecture, so I shall focus on it. But, naturally, he appeals to points he made in the previous lectures, and he will also return to the topic in some of the next lectures; consequently, the interpretive proposals I am trying to articulate here will draw on them as well. Other sources I will rely on are his *Philosophical Investigations* (PI; first published in 1953), and his collection of notes *Remarks on the Foundation of Mathematics* (RFM) edited, and published in 1956, by G.E.M. Anscombe, R. Rhees and G.H. von Wright (the latter was in the LFM audience, as were Casimir Lewy and John Wisdom.)

The structure of the paper is as follows. The next section shall clear the ground for the discussion of what I said I take to be Wittgenstein's main object of analysis when investigating the logicist definition of number in lecture XVI, the relation of one-to-one correlation. Thus, this section will offer a general perspective on what Wittgenstein intends to (and doesn't intend to) concentrate in his examination, while also providing some of the needed background for the so-called Frege-Russell definition of number. (I will not go into more details than I need for the discussion later on; I presuppose familiarity with the basic logicist notions and arguments, as well as with the basics of Wittgenstein's later philosophy). Section 3, the central one, takes up Wittgenstein's issues with the one-to-one correlation idea. The key to understanding what is going on in the discussions in the lectures is, I propose, the notion of *projection*, which I shall explain in the context of his urge (LFM, p. 157) to conceive of number, and counting, in analogy with measuring.[3] It is by way of directing our attention toward these analogies that notions such as *rule-following* and *regularity* (of behavior) enter the scene. Section 4 concludes the paper by tying some lose ends.

2.

It is useful to begin by getting clear on two general directions from which the Wittgenstein of LFM approaches logicism in general, and the definition of number in particular. Answering a question asked by Turing, Wittgenstein clarifies:

[1]There are considerations on the notions of *proof* and *contradiction*, on the status of mathematical propositions, on the analogy between calculations and experiments, and on the role of the logical formalism in relation to natural language—among other things.

[2]At the beginning of lecture XVII, he will introduce another battery of arguments concerning this very notion, but here I have to leave them out as well. I mention them very briefly in Sect. 4.

[3]The term can be found in Wittgenstein's works; he talks about 'methods of projection' in PI (§139, §141, etc.), among other places.

> The point I'm driving at is that Frege and Russell's logic is not the basis for arithmetic anyway – contradiction or no contradiction. (LFM, p. 228)

As this quote suggests, together with the previous one, in LFM Wittgenstein will not only (i) gloss over the differences between Frege's and Russell's definitions of number, but also (ii) will ignore the fact that Frege's system is inconsistent. These two remarks indicate that Wittgenstein is after something transcending the differences between his two great teachers, and orthogonal to the inconsistency problem—which Russell's theory of types manages to solve anyway. Neither will he follow those who object to Russell's reliance on the Axiom of Infinity; also, there is nothing in the LFM on the relevance of Goedel's incompleteness proofs for logicism. Wittgenstein's aim is rather to identify and examine some of the presuppositions supporting the line of thought *shared* by both Frege and Russell, namely the very strategy to define number on the basis of the notion of one-to-one correlation.

He gives the definition in an informal way on p. 156:

> The definition: "A number is a class of classes similar to a given class". "'Similar to' means (we will say for the moment) one to one correlated."

Before we move on, and in the benefit of clarity, we should briefly recall the background of this definition. (The so-called 'Frege-Russell definition of number'.) In *Grundlagen* §68 (Frege 1884/1974), Frege gives the following definition (in modern notation)[4]:

$$\mathrm{NxFx} = \{\mathrm{G}: \mathrm{G} \approx \mathrm{F}\}$$

The expression to define on the left side, 'NxFx', is 'the number of concept F'; it is intended as a (proper) name, and is a singular term denoting an 'object'. On the right side, the expression {x: Kx} is a class-abstraction operator forming the value-range, or extension, of a given concept (K). The elements of this class are the things which fall under the concept (K).

The expression 'G ≈ F' means that the concepts G and F are 'equinumerous', i.e., the elements constituting their extensions can be one-to-one correlated. Thus, the right side says, in essence, that the number of F is a set: the set of those concepts equinumerous with concept F; or, more generally, the class whose members are concepts similar in a certain respect (equinumerosity) with concept F.

More specifically, Frege defines number 0 as the number of the concept 'x ≠ x' ('not identical with itself'), or the set of all concepts equinumerous with 'x ≠ x'. The extension of 'x ≠ x' is of course the empty set Ø, hence 0 is, finally, defined as the set of all concepts with empty extensions. (In this definition, the concept 'x ≠ x' served as a kind of standard, or sample concept.) Further on, number 1 is defined as the number of the concept 'x = 0', number 2 as the number of 'x = 0 or x = 1', and so on (here I gloss over some other technical issues related to the definition of 'successor').

[4]Here I follow the presentation, and the formalism, in Demopoulos and Clark (2005). See also Potter (2000, Chaps. 2, 4, 5).

Similarly, in his *Introduction to Mathematical Philosophy* (1919; Chaps. 2–3) Russell defines, for a class C

(#) The number of C = the class of all classes equinumerous with C.

Thus, 0 is the number of the class with no members, hence 0 is the class of all classes equinumerous with the class with no members; it is this null class which now serves as the sample, or standard class.

Frege's and Russell's definitions of course differ, and this is seen already in the definition of 0; for Russell, 0, and any number, is a *typed* entity.[5] (More precisely, it is a type 2 class which has exactly one member, the type 1 empty set; but 0 is also a type 3 class: the class of all type 2 classes with no members).[6] His definitions are given within his theory of types, which he developed in order to deal with the paradox he discovered in Frege's system. As is well-known, Frege connects extensions with concepts explicitly in *Grundgesetze* **I**, §20 (Frege 1893; 1903/1964), in what seemed an innocuous way, via the Basic Law V (modern notation):

$$\{x:\ Kx\} = \{x:\ Mx\} \equiv (\forall x)(Kx \equiv Mx)$$

But, as Russell pointed out to him in the famous 1902 letter, disaster awaits two steps away. Given Frege's definition of the membership relation $\in$ in second-order logic in terms of the class-abstraction operator (*Grundgesetze* **I**, §34),

$$x \in y \equiv \exists K\ (Kx \,\&\, y = \{x:\ Kx\}),$$

this definition, together with BLV, implies the Naive Comprehension Axiom,[7] hence first-order instances of the Naive Comprehension Scheme, one of them asserting

$$(\exists w)(\forall x)(x \in w \equiv x \notin x),$$

[5]Demopoulos and Clark (2005, pp. 135–6) summarize the difference as follows: "The characterization "Frege–Russell" slurs over the fact that for Russell the number associated with a set (concept of first level) is an entity of higher type than the set itself. Beginning with individuals—entities of lowest type—we proceed first to concepts or sets of individuals and then to classes of such sets (corresponding to Frege's concepts of second level). For Russell, numbers, being classes, are of higher type than sets. But for Frege, extensions, and therefore numbers, belong to the totality of objects *whatever the level of concept with which they are associated.* Thus, while Russell and Frege both subscribe to some version of Hume's principle, their conceptions of the logical form of the cardinality operator—and, therefore, of the principle itself—are quite different: the operator is type-*raising* for Russell, and type-*lowering* for Frege. This difference is fundamental, since it enables Frege to establish—on the basis of Hume's principle—those of the Peano–Dedekind axioms of arithmetic which assert that the system of natural numbers is Dedekind infinite. By contrast, when the cardinality operator is type-raising, Hume's principle is rather weak, allowing for models of it of every finite power."

[6]For (both versions of) the theory of types, see Urquhart (2003).

[7]A proof is in Demopoulos and Clark (2005, p. 132, Footnote 4).

or the existence of a class w consisting of all and only those sets which are not members of themselves. The paradox is derived immediately: $w \in w \equiv w \notin w$.

It is definition (#) that Wittgenstein had in mind while lecturing.[8] As I already pointed out, he set aside the above-mentioned lethal problem (for Frege's system) to which this definition leads. Thus, Wittgenstein is even seemingly willing to grant Frege his magnificent technical achievement, namely the derivation, from the definition $NxFx = \{G: G \approx F\}$, within second-order logic, of the equivalence $NxFx = NxGx \equiv F \approx G$ ('Hume's Principle') and then, via (what is now called) Frege's Theorem, of basic arithmetic.

These technical issues aside, what *does* worry Wittgenstein then? In a rather typical manner for LFM, he says it loud and clear – right after the line where he sketches what he takes to be the logicist definition (#):

> The problem recurs: what do we call one-one correlating […]? We might call it anything.[9] (LFM, p. 156)

Once again, Wittgenstein is not bothered that the expression for one-to-one correlation ($G \approx F$) may be somewhat unclear from a formal point of view. As he knew very well, in *Grundlagen* §71–72 Frege expressed it in second-order logic; yet presenting the formalism is not going to satisfy him. Talking about the formal expression of (#), he says:

> You have substituted another expression for number. – This is the worst of all logical superstitions: the idea that if you write "There is an x such that" and the like, you can solve difficulties of this kind. (LFM, p. 156)

His subsequent analysis, however, is not an attempt to totally reject the logicist formalism, but to expose its limitations:

> I don't want to run down Russell's definition. Although it does not do all of what it was supposed to do, it does *some* of it. (LFM, p. 156)

Now we have to see what, exactly, the Frege-Russell definition (#) is not able to "do".

---

[8](#) is the definition that interests us here, although Russell (and Frege before him) will define, on its basis, number in general: '*A number is anything which is the number of some class.*' (1919, p. 20).

[9]The complete quote is "The problem recurs: what do we call one-one correlating the roots with so-and-so? We might call it anything." Wittgenstein already began discussing the problems encountered when trying to count the roots of quadratic equations, especially those of the type $x^2 - 6x + 9 = 0$—which are said to have two roots, although +3 is the only value satisfying the equality. I left out the reference to the roots in the quote, but we'll discuss them in a moment.

3.

The worry, let's recall, is that what we actually mean by a one-to-one correlation—the gist of the Frege-Russell definition—is, despite strong appearances to the contrary, left in the air ("We might call it anything.") In fact, Wittgenstein introduces this worry earlier on in the lecture XVI by puzzling over a seemingly trivial question: how many roots does a quadratic equation have?[10] The answer is, supposedly, given by first finding the roots, and then counting them. But, as I have already suggested in Footnote 9 above, this problem presents interest because there is a type of quadratic equation for which only one numerical value satisfies it—and we count this solution twice. (An example is $x^2 - 6x + 9 = 0$.)

The problem here is quite easy to understand. If we are explained what a root is—the numerical value which, when uniformly substituted for the variable, yields the equality– then there is an obvious answer to the question regarding the roots of $x^2 - 6x + 9 = 0$: it has one root (+3). In other words, if we actually count the solutions in the usual manner, by one-to-one correlating the elements of the set of solutions S with the natural numbers, we get as far as one. However, like any other quadratic equation, ours is said to have two roots. This means that, somehow, if we one-to-one correlate S with the natural numbers we get as far as *two*. What is going on?

In the lectures Wittgenstein doesn't get into any details as to why we count the same value twice. But, as mathematicians will likely tell us, the requirement to have two solutions is introduced because one can prove a mathematical proposition called 'the fundamental theorem of algebra', which states that every degree *n* polynomial with complex coefficients has exactly *n* roots. This theorem, as its name suggests, is one of the deepest theorems in all mathematics, absolutely pivotal in connecting various strands of traditional arithmetic (concerned with solving equations) with more advanced branches of algebra, such as the theory of complex numbers and the theory of fields (in fact, an equivalent statement of the theorem is that the field of complex numbers is algebraically closed.).[11]

The role of the roots example is to draw attention to the ignored complexities of the notion of counting (and of one-to-one correlation), which originate in the fact that counting is never *just* counting. To say that we do it by one-to-one correlating natural numbers with objects is insufficient, since this operation always comes embedded into a context, which in turn contributes to determining the result (and this can be quite different from what we may expect.) This is the sense of Wittgenstein's remark later on in the lecture when he returns to the roots example:

[10]This is an example of the use of the notion of counting *inside* mathematics, a use that interested Wittgenstein to the highest degree, especially in relation to Cantor's theories (more on this below).

[11]As Mark Steiner pointed out to me (personal communication) Wittgenstein had probably more elementary reasons in mind. If we consider the general form of a quadratic equation $x^2 + bx + c = 0$, the product of the roots is c, and the sum of the roots is –b, and for these relations to hold we have to assume two roots.

> That we know how to count is presupposed in Russell's definition of number. (LFM, p. 160)

In saying this, Wittgenstein's intention is not to raise a direct objection; the point of the remark is *not* that the definition is circular in a somewhat trivial way.[12] Against the background I sketched so far, I suggest that we should read it as saying: 'that we know *what to call a one-to-one correlation* is presupposed in Russell's definition'. Russell would agree with this, and perhaps won't see any problem. Yet Wittgenstein would insist that we *don't* know whether to call something a one-to-one correlation until the context is given. The notion of one-to-one correlation is context-sensitive[13]; as we just saw, the set of solutions of *any* quadratic equation has two members in the context of the general algebraic theory. Wittgenstein continues by emphasizing the general point:

> [T]he definition doesn't *fix* (…) what in this case [the roots puzzle] we are going to call one-one correlation. (LFM, p. 160)

In fact, these remarks echo the pronouncements opening lecture XVI:

> We are inclined to say, 'We count in order to find the number of things.' Compare: 'We weigh in order to get the weight.' Both statements are in a way equally fishy. (LFM, p. 152)

How are they fishy? While it makes sense to say that we count (a set of things) to find how many there are, to say that this is why we count them is misleadingly insufficient. We count to find the number of things, sure enough, but then we normally are going to *do* something with the number—and this is why we count, because we need the number-result for some further use, depending on the context in which we count (e.g., when we have people over for dinner, we count the plates on the table not just to find out that there are five of them, but, for example, to make sure that each of our guests will have one.).

Back to the roots issue, we see that in saying that our equation has two solutions we have in fact made a *decision*—and, we should add, a well-justified one, given the context of the general algebraic theory in which the question is embedded. I submit that this is how we should read Wittgenstein's provocative remark that

> You've got to decide what you're going to call a one to one correlation with so-and-so. (LFM, p. 156)

---

[12]The (few) interpreters interested in later Wittgenstein's take on the logicist definition of number focus on this circularity (*petitio*) charge. Marion (1998, pp. 77–83), for instance, discusses a version of the objection that reached us via Waismann (1979). Such a charge was in fact made by Poincare, and Goldfarb (1988) argues that it is flawed (see Folina (2006) for discussion.) I am of the view that (i) Goldfarb is right, and (ii) Wittgenstein did not have in mind the Poincare-Waismann objection. De Bruin (2008) comments on these issues, but I can't tell if he sees Wittgenstein's objection as different in essence from Poincare's.

[13]See also LFM, p. 161: "The question is: What is one to call a one-one correlation? You have the example of the cups and saucers; and then you think you know under all circumstances what the criterion of one-one correlation is." I'll comment on this later.

(Moreover, we now also see in what sense he says that "we might call [the correlation] anything": the decision depends on the context, and this can be 'anything'.)

But, interestingly, let's ask if we can insist, after being told about the context and the bigger picture, that the equation has *one* root. Does it seem acceptable to *decide otherwise,* namely to ignore the larger mathematical context to which the question (and the algebraic expression '$x^2 - 6x + 9 = 0$') belongs? Thus, are we entitled to retort to the more sophisticated mathematician that it was us, the simple-minded, and not the sophisticated mathematician, who actually answered the posed question—since we did *the same thing* we always do when the 'how-many?' question is asked, namely just counted the roots.[14] This is perhaps so, and thus it seems acceptable to hold fast to the 'one root' answer. Such a decision, however, would render us mathematically deviant: we would not be able to recognize the fundamental theorem of algebra because we found a counterexample to it!

As I announced at the outset, I take Wittgenstein's endorsement of the analogy between counting and measuring to constitute the most important clue in understanding his take on the notion of one-to-one correlation (and thus on the Frege-Russell definition). Here is the relevant paragraph:

> We must always think of *number* as we think of *length* or *weight*, and of *counting* or *correlating* as we think of *weighing* or *measuring*. (LFM, p. 157)

In addition to presenting the advantage that the proposal is explicitly authorized by Wittgenstein himself (so we don't have to wonder *who is talking* now, situation causing lots of troubles in the PI[15]), the idea expressed here deserves thorough analysis because, despite its opaqueness, holds the promise to open several helpful interpretive doors.

To begin with, let's note that Wittgenstein's proposal seems a bit odd: how can *number* be put on the same footing, so to speak, with *length* (or weight), and counting with measuring? They seem to belong to different planes, so to speak. Lengths are typically given *as* numbers; numbers are one thing (a fundamental notion, we think), lengths another (a derived notion: expressed as a number of units), so how can it be illuminating to mix them up—and suggest to think of the former by analogy with the latter?

What may help us grasp what Wittgenstein is driving at by this proposal is, I suggest, the introduction of the more general notion of *projection*. What we do when we count the things in a set is to (mentally) project the things onto the series of natural numbers; what we do when we measure the length of a stick is to (literally) project its ends onto a ruler—such that we one-to-one correlate the things

[14]One recognizes here the central Wittgensteinean theme from his discussion of rule-following in the PI, that 'the same' is meaningful only when embedded into a context, or a practice. More on this below.

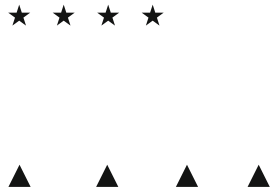
[15]Stern (2004) draws attention that different 'voices' take different positions in the PI, and thus any interpretation has to sort out who is saying what, and especially what Wittgenstein himself endorses from what is being said.

with numbers, and the ends with two marks on the ruler. Counting, measuring (lengths), one-to-one correlations in general, are thus *projective operations* (or *techniques,* as Wittgenstein often calls them.) The numerical values we get in the end are the length of the stick and the 'length' of the set. Thus, to one-to-one correlate is to use certain projective techniques.

What does Wittgenstein achieve by introducing the analogy counting ∼ measuring? An entirely new perspective on the matter, since, as is well known, there are various ways one can measure the same quantity—and thus the further key-question arises right away: do these various ways have to give the same result? Equivalently, there are various kinds of projections, and, as we also know very well, different projective operations result in different projections. We'll return to these points later. Now I will go over a series of examples, in the hope that they will make these remarks about the role of projection clearer.[16] I shall begin with a simple one, which will serve as basis for a longer discussion, and finish with an illustration involving the one-to-one correlation of infinite sets. (Wittgenstein comments on this issue in LFM, pp. 160–161.)

It is a well-known psychological fact that children up to a certain small age have troubles with answering questions about equinumerosities. If they are presented with two rows of objects arranged as follows (Fig. 1),

they will answer that there are as many stars as triangles. But, if the arrangement is changed into this one (Fig. 2),



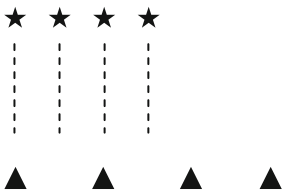
[16]None of them is Wittgenstein's, as far as I can tell, but they are occasionally inspired by his own.

they will say that there are more triangles than stars. Thus, they systematically fail this 'number-conservation task' (as called by Jean Piaget, who first studied it). The wrong answer result is experimentally robust—and stunning; psychologists report that parents can't believe that their children gave it, and plan to repeat the test at home.[17] There can be no doubt that anyone who gives the wrong answer doesn't (yet) understand the *meaning* of the expression 'as many as'.

The error is a projection error. In the case of the first display (Fig. 1), pre-arithmetical children use what seems a natural parallel projection of one set of objects onto the other set. To get the right answer, they probably do this (Fig. 3) or, perhaps, this:

However, when the row of triangles gets dispersed, the use of the same type of parallel projection is a possible explanation as to why they get the wrong result. When they report that there are more triangles, they may do this:

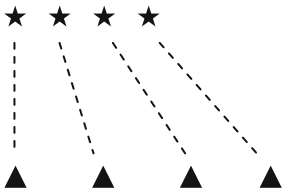
[17]Gelman and Gallistel (1986, p. 1). They analyze more such tasks in Chap. 1.



What is going on in children's minds is difficult to say, and is the psychologists' job to tell us; it will be immaterial for what follows. (How and why, exactly, do they get the wrong answer: do they project like in Fig. 5, or in the manner of the diagram in Fig. 4, or maybe they do something entirely different?) However, what will be significant later on is another robust psychological fact reported by the experimenters: namely, how amazingly *fast* the children 'get it right' after being explained what they are supposed to do, and how little they err afterwards.[18]

The scenario to discuss now is the situation in which a child—call her Deviant Sara—dares to voice her dissent, claiming that she did *not* get the wrong result. Sara's contention is that there are *more* triangles than stars in the dispersed display in Fig. 5, because she got this result by doing *the same thing* she did the first time (see Fig. 3)—and then we agreed that the result was right.

Our immediate reaction here, I suspect, is to explain to Sara that, despite certain similarities, the second time (Fig. 5) she *didn't* do the same thing: the similarities between the projective operations illustrated in Figs. 3 and 5 are superficial. What she did the same the second time (namely, ensuring that the projecting lines are parallel) was *irrelevant* for the question we asked, while the truly relevant aspects of the procedure are surely not the same the second time. This is in essence to point out to the deviant child that she should not have used the same type of projection (i.e., a parallel one), but a different type, such as the one shown below:



[18]See Gallistel and Gelman (1986, Chap. 2), especially the learning acquisition curve on p. 14. It should be noted that their study aims to argue for *nativism*, the view that such capacities are innate. I will not comment further on this thesis, but even if it were true, the relevance of this point for the discussion here would be unchanged. I discuss some possible connections between more recent proponents of nativism (Wynn, Spelke, Carey, etc.) and later Wittgenstein's philosophy of mathematics in Bangu (2012a).

(And we can of course think of other correlations; an unusual one will be discussed below.)

To reply to Deviant Sara in this way is to stress the modal character of the procedure: a numerical equality (of two sets of objects) is established when such one-to-one correlation *can* be demonstrated. This amounts to advising her that she should not get stuck to using only one type of projection (the parallel one), but should try a variety of projections: if she is able to find one *of the right kind*, then she will answer the question. But this is not the end of the matter, since she will insist that we specify what 'the right kind' means. (This—as is perhaps already clear—is exactly the question Wittgenstein began with, in another guise: what do we call a one-one correlation, after all?)

The above 'can', although it may sound natural, leads to even deeper difficulties. Wittgenstein's discussion in lecture XVI follows (unsurprisingly) exactly this path:

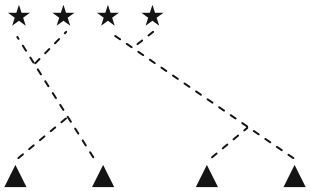
> "Similarity is decided by one-one correlation." – "Being correlat*ed* to" – or "being *able* to be correlated"? (Compare: the possibility of drawing lines.) (…)
>
> Suppose I said, "We will define all numbers by means of one-one correlation". Russell says that a class has this number if it can be correlated one-one to a standard class. This is like "A thing is a foot long if it can be made to coincide with the Greenwich foot." Is it always possible to bring it to the Greenwich foot? (…) Suppose it is impossible to carry this out. (…)
>
> "It is possible for it to coincide" – doesn't it depend on what way? We might say: if we heat or stretch this, it coincides with the Greenwich foot. Now this must either be forbidden or allowed. (LFM, pp. 157–158; italics in original.)

An intermediary conclusion is drawn a few lines down the page:

> If I have left out the way of making this coincide with the Greenwich foot, thereby I've left out everything. (LFM, p. 158)

This remark repeats the point we discussed above (that "we might call it anything"), but here puts it in terms of measuring. If we are to re-formulate it in terms of counting (i.e. 'measuring' sets of objects) and projecting (one-to-one correlating), the thought would be that the notion of one-to-one correlation is empty until we specify which projections are 'of the right kind', i.e., which are allowed, and which are forbidden.

At this point, even those vaguely familiar with later Wittgenstein's way of thinking will not be surprised to hear that such a specification is virtually impossible to formulate in a general and all-encompassing fashion; the Wittgensteinean theme to which I allude here is the so-called 'rule-following paradox' ('the paradox of interpretation') presented in PI §201: "any course of action can be made out to accord with the rule", hence "no course of action could be determined by a rule." A simple example will hopefully suffice to show the difficulty. Suppose that Sara is asked the question again (are there as many stars as triangles?), and this time she answers the question (correctly) by using this projection:

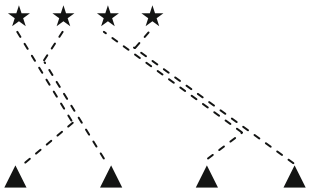


Is this allowed or forbidden? Can we accept it as a one-one correlation? Note that Sara uses it to provide the right answer—yet, on pain of circularity, we have to exclude this criterion in assessing it. But, then, what *are* the criteria? In fact, Wittgenstein warns about this himself later on, at the beginning of lecture XVII:

> The question is: What is one to call a one-one correlation? You have the example of the cups and saucers; and then you think you know under all circumstances what the criterion of one-one correlation is. (LFM, p. 161)

In Fig. 7 we face one of these unusual "circumstances". By using this projection, does Sara follow the order 'one-one correlate the elements of the two sets of objects'? It is unclear what we should say. One may be inclined to say 'no': for one thing, unlike the typical cases of correct, typical projections (see Figs. 3 and 6), here the correlation is inherently ambiguous, since one can't indicate which triangle is correlated to which star (and vice versa).

I mentioned above the rule-following paradox, and we can now see that this situation is an illustration of it. So, once again, did Sara obey the order to one-one correlate the stars and the triangles? Contrary to the negative answer above, we can also say that she actually did obey it, because—recall: 'any course of action can be made out to accord with the rule'—we can make her correlation out to accord with the rule (as *we* understand it). For instance, take the connecting lines to stand for *superimposed lines*, each of them linking a unique star and a unique triangle, as follows (Fig. 8):



(In fact, doing this is not as mad as it seems, since it provides a possible explanation as to how Sara managed to get the right answer despite the ambiguity!) The point of these considerations is quite straightforward: it can be a difficult matter

to assess whether or not certain projections qualify as one-to-one correlations. There are easy cases either way, but there are intermediary cases as well: we simply *don't* always know whether a certain projection qualifies as a one-one correlation. We'll return to this problem below, where it will appear infinitely more difficult (pun intended), since we'll discuss Wittgenstein's thoughts on the one-to-one correlations between infinite sets.

At this stage in our discussion we have to remember the lesson that the paradox of rule-following is meant to teach us. The point of reflecting on the paradoxical situation is *not* to draw the skeptical conclusion that following a rule, or an order—the order here is: 'one-one correlate so-and-so!'—is impossible,[19] in the sense that anything I do is, after all, acceptable—and thus I need something (mental), or somebody (the larger community) to keep me from erring. Rather, Wittgenstein concludes that the paradox shows that "there is a way of grasping a rule which is *not* an *interpretation*" (PI §201). I take it that the very point of the rule-following story is to bring to the forefront this otherwise well-known, but never fully appreciated *brute fact*: that we humans *can*, and *do* obey rules, all the time. As Fogelin put it: "From an abstract point of view, anything can be shown to be in conformity with the instructions we have given [someone]. Yet people do, on the whole, follow such instructions correctly, so again (…) a conceptual indeterminacy overbalanced by nothing more than a brute fact of human nature." (Fogelin 2009, p. 99)[20]

This fact of human nature appears in our discussion of arithmetic in the form of a regularity of behavior: when asked to establish one-one correlations (i.e., to pair objects) in usual circumstances, the vast majority of minimally educated adults will immediately think of using projections like the ones illustrated in Figs. 3 and 6. Exceptions exist, no doubt, and moreover, some people—the idiots, in the medical sense of the term—will fail to grasp the technique at all. But were failures (of Fig. 5-type) or confusions (of Fig. 7-type) widely-spread, were we constantly quarreling over whether to call a certain projection a one-one correlation, the arithmetic as we know it now would be impossible.[21] On the whole, for virtually all of us (including very small children), after only very little instruction, following this order poses no difficulties. For Wittgenstein, this is a fact (a statistical regularity, but a very robust one), immediately verifiable, uncontroversial, and yet of outmost importance. It is also an unremarkable and familiar fact, always under our eyes—hence we pay no attention to it. For this reason it is one of those aspects of our life which Wittgenstein aims to make visible to us, to *describe* to us as the rough ground (PI §107) supporting our practices; in this case, it makes possible the arithmetical practice.

[19]Needless to say, there is no consensus on what is the right interpretation of these passages.

[20]Fogelin points out here the resemblance between Wittgenstein's and Hume's views. See also Fogelin (1987, p. 216).

[21]Here it should be added that the world itself also cooperates in making arithmetical practice possible, since virtually all usual objects—pebbles, pens, fingers, ink-marks, chairs, people, cars, etc.—subsist long enough, and don't divide or multiply fast enough, to be counted by creatures like us.

The existence of regularities of behavior like this—you ask people to pair objects, then they will do (roughly) the same thing, end of story—is, to use Wittgenstein's own words, "an immensely important fact about us human beings." (LFM, p. 182) This is so because these regularities ground not only arithmetic, but language itself. The point is made in PI §207, and also in LFM in lecture XIX:

> For instance: you say to someone "This is red" (pointing); then you tell him "Fetch me a red book" – and he will behave in a particular way. This is an immensely important fact about us human beings. And it goes together with all sorts of other facts of equal importance, like the fact that in all the languages we know, the meanings of words don't change with the days of the week.
>
> Another such fact is that pointing is used and understood in a particular way – that people react to it in a particular way.
>
> If you have learned a technique of language, and I point to this coat and say to you, "The tailors now call this colour 'Boo' ", then you will buy me a coat of this colour, fetch one, etc. The point is that one only has to point to something and say, "This is so-and-so", and everyone who has been through a certain preliminary training will react in the same way. We could imagine this not to happen. If I just say, "This is called 'Boo'" you might not know what I mean; but in fact you would all of you automatically follow certain rules.
>
> Ought we to say that you would follow the *right* rules? – that you would know the meaning of "boo"? No, clearly not. For which meaning? Are there not 10,000 meanings which "boo" might now have? – It sounds as if your learning how to use it were different from your knowing its meaning. *But the point is that we all make the SAME use of it.* To know its meaning is to use it *in the same way* as other people do. "In the right way" means nothing.
>
> You might say, "Isn't there something else, too? Something besides the agreement? Isn't there a *more natural* and a *less natural* way of behaving? Or even a right and a wrong meaning?" – Suppose the word "colour" used as it is now in English. "Boo" is a new word. But then we are told, "This colour is called 'boo'", and then everyone uses it for a shape. Could I then say, "That's not the straight way of using it"? I should certainly say they behaved unnaturally. (LFM, p. 182-183)

Returning to our main concerns, we have seen that according to Wittgenstein it is precisely this fact, the existence of *agreement when it comes to judging what qualifies as one-one correlation* (i.e., the convergence on the method of comparison) that gives the definition of number its strong conceptual appeal—and *not* its formulation in logical terms:

> The definition [#] seemed illuminating because we immediately think of one method of comparison. (LFM, p. 158)

Before we move on to discussing other examples of projections (this time in the infinite domain), let's note that talking in terms of 'agreement' is potentially misleading. Many (some illustrious) readers of later Wittgenstein's works interpret him as proposing a version of conventionalism about mathematics, and thus as doubting the objectivity of mathematics: since he stresses the importance of agreement among us in matters arithmetical, then he must be decreeing that the right result (of counting, or some other operation) is what we (the community of calculators) take

to be the right result—hence Dummett's 'full-blooded conventionalist' label.[22] Yet this is surely a misunderstanding. The lucid Wittgenstein exegesis spotted the mistake,[23] and explained it quite clearly: for Wittgenstein, the agreement among the *opinions* voiced by people on arithmetical propositions doesn't establish their truth-value. Gerrard (1996), for one,[24] gathers a wealth of textual evidence to support this point. (I draw on this evidence below). Thus, we read things like these:

> Certainly the propositions 'Human beings believe that twice two is four' and 'Twice two is four' do not mean the same. (PI II, xi, p. 226)
>
> Mathematical truth isn't established by their all agreeing that it's true (LFM, p. 107)
>
> It has often been put in the form of an assertion that the truths of logic are determined by a consensus of opinions. Is this what I am saying? No. (LFM, p. 183)

Truth-values in arithmetic are not established by our agreeing on the results—and yet agreement does have a fundamental role to play.[25] How can this be? To understand this, we need to pay attention to Wittgenstein's distinction between two kinds of agreement. PI §241 separates "agreement in opinions" from agreement in "form of life" (or, as Wittgenstein also says in PI §242, agreement in "judgments"). On the one hand, there is agreement in verbal, discursive behavior, or in "*opinions*" (LFM, p. 183). But, on the other hand, there is that other kind of agreement, of the behavioral type—the consensus "of *action*" (LFM, p. 183)—we discussed above. This is the relevant kind, the agreement in "what [people] do" (LFM, p. 107).

> What is it we agree to? Do we agree to the mathematical proposition or do we agree in *getting* this result? These are entirely different.
>
> What is it they must agree in? They agree in *getting* this. They may agree in saying "I got so-and-so" – in finishing up with the same number, etc. But not that this is the answer. (…) They agree in what they do. (LFM, p. 107)

Moreover, In RFM, VI-30 Wittgenstein points out:

> The agreement of people in calculation is not an agreement in opinions or convictions.

He also insists:

> There is no *opinion* at all; it is not a question of *opinion*. They are determined by a consensus of *action*: a consensus of doing the same thing, reacting in the same way. There

[22]On Dummett's (1959) reading, Wittgenstein goes as far as claiming that at any step in a calculation we could do what we want. Later on, he notes: "What makes a […mathematical] answer correct is that we are able to agree in acknowledging it as correct." (1978, p. 67) Cited in Gerrard (1996, p. 197, Footnote 43).

[23]The next several paragraphs draw on sections in Bangu (2012b).

[24]Fogelin (1987) also titles the first section in his chapter on Wittgenstein's philosophy of mathematics 'Anti-Platonism without conventionalism.'

[25]Wittgenstein is of course aware of how his point can be misunderstood. He notes: "It has been said 'It's a question of general consensus'. *There is something true in this*." (LFM, p. 107; my italics) And then goes on to clarify what kind of consensus, or agreement he has in mind.

> is a consensus but it is not a consensus of opinion. We all act the same way, walk the same way, count the same way. In counting we do not express opinions at all. (LFM, pp. 183–4)

As we saw, behavioral agreement (i.e., in action) is a precondition of the existence of arithmetical practice.[26] This agreement has, one may say, a *constitutive* role for the arithmetical practice; it must already be in place before we can even speak of 'mathematics'.[27]

Having introduced and discussed the role of *projection* (as a generalization of the notions of counting and measuring) and *agreement* (in using certain projective techniques), we can now examine the last point Wittgenstein made in lecture XVI about the notion of one-to-one correlation. On the last two pages of the lecture, we read:

> It is said to be a consequence of Russell's theory that there are as many even numbers as cardinal numbers, because to every cardinal number I can correlate an even number.
>
> But suppose I say, "Well, go on – correlate them." Is it at once clear what I mean? (LFM, pp. 160–161)

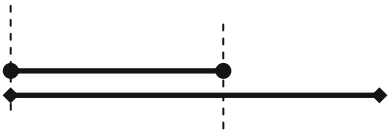
The question is rhetorical, and the answer is 'no'. I'll close the paper with some remarks on what Wittgenstein may have had in mind here.

Heeding his advice, let's return to the analogy between counting and measuring. We'll examine the following situation. Suppose we point to Deviant Joe a pile of sticks and ask him to bring us a stick of the same length as our stick *s*. He brings us stick *j*, below.

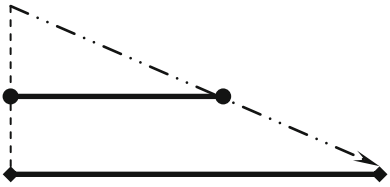
We are puzzled; we superimpose the two sticks and show to him that stick *j* is not as long as *s*:

[26]Hacker (2010) explains: "The agreement of human beings that is presupposed by logic is not an agreement in opinions (RFM 353). Similarly, the agreement of people in calculation is not an agreement in convictions (RFM 332). It is an agreement in form of life, and that means: an agreement in concepts and their application, and hence in the behaviour consequent upon their application. It is an agreement on the measures by which we judge reality, and hence also, an extensive agreement on the results of measurement (PI §§241–2)."

[27]These empirical regularities are 'hardened' into rules (RFM VI-22), and this explains why the later Wittgenstein links systematically the sense of mathematical propositions with their application: mathematical propositions are "dependent on experience but made independent of it" (LFM, p. 56), and thus have a *normative* character (unlike the empirical propositions). For more on this, see Steiner (1996, 2009); for later Wittgenstein's views on the sense of mathematical propositions, see Rodych (2008).



Deviant Joe, of course, disagrees. His way of comparing sticks assures him that *j* and *s* are of equal length. He explains his procedure to us. See Fig. 10. He begins by doing what we did in Fig. 9, and finds which stick is 'shmorter' (We will call it 'shorter'.) Then, he raises a perpendicular from one end of the other stick (*s*, in this case), and then positions the shmorter stick (*j*) parallel to *s* at a point on the perpendicular which he determines by stepping next to it—we would say that it corresponds to the length of his foot. He then extends that perpendicular a distance equal to his foot, once again. Then he connects the end of this perpendicular to the other end of *j*, and then extends this connector line until it reaches (or not) stick *s*. In fact, three situations are possible, and he answers the comparison questions depending on which one obtains. First, it may happen that the connector line reaches beyond *s*, and then Joe says that *j* is 'longer'. Second, the connector line may reach exactly the end of *s*, and he'll say that *j* is 'as long as' *s*. Third, the line may reach within *s*, and then Joe will say that *j* is 'shorter' than *s*. As is clear, stick *j* is the sh(m)orter one; the two vertical dotted segments in Fig. 10 are equal, and equal to Joe's foot. As the figure indicates, he has to conclude that *j* is as long as *s*.



How are we to react to Deviant Joe's claim? We can give him several reasons to prefer our parallel projective technique (Fig. 9) to fix the meaning of 'as long as' (and 'longer than'). For one thing, his 'longer than' is not transitive, and his 'as long as' is neither symmetrical, nor reflexive. But he may well reply that his method of comparison *doesn't have to have* these features (which our method has); it is good enough for him that his method will always give an unequivocal answer to any comparative question. Moreover, he may point out, the methods agree quite often in selecting the longer stick from a pair of sticks (although they always disagree when it comes to equally long rods). Strangely enough, the methods are guaranteed to

agree every time when the longer stick is *much* longer than the other stick. Hence he may insist that his method is *similar enough* to our method: in fact, it is a more robust, and less error-sensitive, *extension* of our method.[28]

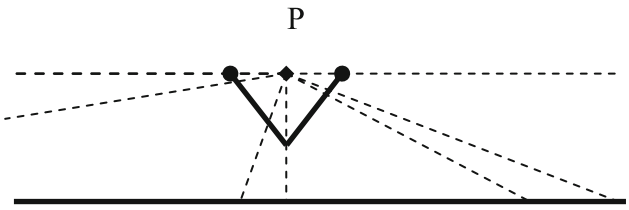
This is all fine, but it doesn't mean that we should accept Joe's claim that he just found a stick as long as ours. He didn't find *that*; he didn't *discover* such a stick in the pile. To present what happened as a 'paradox' (of the sticks!) is at best a joke: what he found was *a method to compare* sticks (a projection) which delivered this result. Or, better put: he *invented* such a method/projection. He didn't find a stick *j* having the property we wanted, but invented a new way of looking at *j*. And this, in the end, is pointless: if Joe himself needs a stick to reach an apple on a remote branch of a tall apple-tree, he'll ask for stick *s*, not *j*—because stick s *is* longer, in the only sense in which 'longer' genuinely has a meaning (i.e., use).

Having learned this lesson, we can now return to Wittgenstein's complaint that 'it isn't clear' what is going on when we (mathematicians) say, following Cantor, that (paradoxically) the natural numbers are as many as the even numbers—and we say this because we are seemingly able to one-one correlate them (in the usual fashion: $1 \rightarrow 2$, $2 \rightarrow 4$, etc.) On this, Wittgenstein remarks:

> But you have not found two classes which have the same number; you have only invented a new way of looking at the thing. (LFM, p. 161)

We can now see that there is no difference between what we did in the case of the two sets of numbers and what Deviant Joe did with the sticks: we just invented a new projection. In fact, when it comes to inventing projections, there is virtually no limit to how strange they can be: as is well known, one can even show that any finitely long segment is as long as an infinitely long one! See Fig. 11 (Galileo is credited with inventing this type of 'paradox'.) First, we break *j* in two equal parts, but keep them joined. Then consider a point P midway between the two ends of the cracked stick; we position the ends such that a line drawn through them is parallel to the infinitely long line *l*. Then we say that *j* is as long as an infinite line *l* because we can one-one correlate all the points of *j* with all the points of *l*, as follows. We pick an arbitrary point on *j* and join it with point P, then we continue the joining line until it reaches *l* (it will never fall outside since *l* is infinite). Conversely, pick any point on *l* and join it with P; the connecting line will intersect *j* in exactly one point. We also say that the two end points of *j* correspond to the end points of *l*—which are located 'at infinity', to the left and to the right. It seems that *we now did the same things* we normally do when we show that two sticks have the same length: what is *different*?

[28]Suppose there is a strange kingdom high up in the mountains, where the atmosphere is such that the best approximation of parallelism for two rods obtains when they are one foot apart. Moreover, the priests have decreed that it's a monstrous sin to simultaneously draw perpendiculars from the two ends of a segment—so no one even dares to think of doing this. In this kingdom, Joe's method of comparison may be taken seriously.



In the case of the numbers, Wittgenstein also notes that we 'haven't even yet correlated any two things':

> In fact, if you say you have one-one correlated the even numbers to the cardinal numbers – you have shown us an interesting extension of this idea of one-one correlation. But you haven't even yet correlated any two things. (LFM, p. 161)

It should now be clear how to read this. Normally, i.e., in usual situations, in the relatively small finite domain, we have no difficulties to one-one correlate the elements of two sets (or to compare sticks); there is robust agreement on this, and no paradoxes lurk in the dark. However, when we face infinite sets, we obviously can't project the elements of these sets as we usually do, so now—importantly—we never *actually* correlate the elements. Instead, we tacitly operate with an *extension* of the idea of finite one-one correlation—an extension into the infinite domains. We now one-one correlate* these domains, so to speak. But an extension of an idea is not the very idea: a one-one correlation* is not a one-one correlation, no matter how similar the two strike us to be. And, just as with the sticks, we should normally think that such an extension is no good (it leads to paradoxes after all!), although, again, just as with the sticks, we can imagine circumstances when it is of some use (see Footnote 27).

To conclude, what drives Wittgenstein's thoughts here is *not* his supposed sympathy for finitism.[29] (This can't even be a working hypothesis, since the finitists urged interference with mathematics as practiced, while Wittgenstein vehemently, and programmatically, opposed it. See PI §124–5, among other places.) The aspect he is worried about here is not so much the mathematical infinite,[30] but a certain *attitude* (in particular, Cantor's and Hilbert's) toward a judgement of similarity: once we accept that a certain projection qualifies as a one-one correlation in the finite domain, then we are bound to say that another, *similar* projection also qualifies as a one-one correlation (this 'another projection' is the extension of the initial projection to the infinite domain.) We *can* of course say this, but we *don't have* to; as he pointed out earlier, this is not a given—i.e., a mathematical 'reality' we discover in awe—but a *decision*.

[29]See Frascolla (1994, part 3), and Marion (1998) for discussion. Sect. 2 in Rodych (2011) is also informative.

[30]See Moore (2001, pp. 137–140, 206–207).

4.

Let me close by reissuing my initial warning, that in this paper I am not able to deal with everything that Wittgenstein had to say in the LFM about the Frege-Russell definition of number; what is more disappointing, I can't even cover all he says in the lectures about the notion of one-to-one correlation. More specifically, he continues its investigation in lecture XVII, and the second paragraph reads as follows:

> Russell seems to have shown not only that you *can* correlate any two classes having the same number, but also that any two classes with the same number *are* correlated in this way. This at first sight seems surprising. But he gets over this, as Frege did, by the relation of identity.
> There is one relation which holds between any two things, a and b, and between them only, namely the relation $x = a \,.\, y = b$. (If you substitute anything except a for x or anything except b for y the equations become false and so the logical product is false.)
> You go on for 2-classes:

$$\begin{array}{cc} a\ b & c\ d \\ x = a.\ y = c. & \mathbf{v}\ .\ x = b.\ y = d \end{array}$$

> And so you go on to classes of any number. And so we get to the surprising fact that all classes of equal number are already correlated one-one. (LFM, p. 162; the passage ends with a footnote reading: 'Cf. *Philosophical Grammar* pp. 355-358')

After illustrating how this works (with a hilarious example), he says:

> There is something queer here. But there is a strong temptation to make up such a relation. (LFM, p. 162)

Once again, Wittgenstein will not be exceedingly happy with this strategy. Yet, as I said, explaining why is beyond the scope of this paper. However, it comes perhaps as a relief that the discussion of these issues takes us back to a territory much better covered by some major exegetical voices. When commenting on the issues raised by the above formalism, Wittgenstein will express worries about the limitations of the Frege-Russell 'predicate logic'—and this strand of his thoughts has been already examined in detail. A list of such works include, among others, Baker and Hacker (1984, 1989), Hacker (2001, Chap. 8) and Floyd (2005), and thus I direct the reader to them.

Acknowledgments I thank Simo Saatela, Penelope Maddy and Mark Steiner for reading a first draft of this paper. Needless to say, I'm the only responsible for all remaining infelicities or errors.

References

Baker, G. P., & Hacker, P. M. S. (1984). *Frege: Logical excavations.* Oxford: Oxford University Press.

Baker, G. P., & Hacker, P. M. S. (1989). Frege's anti-psychologism. In Notturno, M. (Ed.) *Perspectives on psychologism*. The Netherlands: Brill.

Bangu, S. (2012a). Wynn's experiments and later Wittgenstein's philosophy of mathematics. *The Jerusalem Philosophical Quarterly, 61*, 219–241. (*Iyyun*).
Bangu, S. (2012b). Later Wittgenstein's philosophy of mathematics. *the internet encyclopedia of philosophy*. http://www.iep.utm.edu/wittmath/
Canfield, J. (1981). Review of Wittgenstein's lectures on the foundations of mathematics. Cambridge 1939. *Canadian Journal of Philosophy, 11*(2), 333–356.
De Bruin, B. (2008). Wittgenstein on circularity in the Frege-Russell definition of cardinal number. *Philosophia Mathematica, 16*(3), 354–373.
Demopoulos, W., & Clark, P. (2005). The logicism of Frege, Dedekind, and Russell (pp. 129–165). Shapiro.
Dummett, M. (1959). Wittgenstein's philosophy of mathematics. *The Philosophical Review*, *LXVIII*, 324–348.
Dummett, M. (1978). Reckonings: Wittgenstein on mathematics. *Encounter*, *L*(3), 63–68 (March 1978).
Floyd, J. (2005). Wittgenstein on philosophy of logic and mathematics (pp. 75–128). Shapiro.
Fogelin, R. J. (1987). *Wittgenstein*. New York: Routledge & Kegan Paul.
Fogelin, R. J. (2009). *Taking Wittgenstein at his word*. Princeton: Princeton University Press.
Folina, J. 2006. 'Poincaré's circularity arguments for mathematical intuition. In M. Friedman & A. Nordmann (Eds.), *The Kantian legacy in nineteenth-century science* (pp. 275–294). Cambridge, Mass: MIT Press.
Frascolla, P. (1994). *Wittgenstein's philosophy of mathematics*. London and New York: Routledge.
Frege, G. (1884/1974). *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl* (Breslau: W. Koebner, Trans.). *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number*, by J.L. Austin, Oxford: Blackwell, second revised edition, 1974.
Frege, G. (1893; 1903/1964). *Grundgesetze der Arithmetik*, Jena: Verlag Hermann Pohle, Band I/II. Partial translation of Vol. I, *The Basic Laws of Arithmetic*, by M. Furth, Berkeley: U. of California Press, 1964.
Gelman, R., & Gallistel, C. (1986). *The child's understanding of number*. Cambridge, MA: Harvard University Press (1978 1st ed.).
Gerrard, S. (1996). A Philosophy of mathematics between two camps. In H. Sluga & D. Stern (Eds.), *The Cambridge companion to Wittgenstein* (pp. 171–197). Cambridge: Cambridge University Press.
Goldfarb, W. (1988). Poincare against the logicists. In W. Aspray & P. Kitcher (Eds.), *History and philosophy of modern mathematics* (pp. 61–81). Minneapolis: University of Minnesota Press.
Hacker, P. M. S. (2001). *Wittgenstein: Connections and controversies*. Oxford: Clarendon Press.
Hacker, P. M. S. (2010). A Normative conception of necessity: Wittgenstein on necessary truths of logic, mathematics and metaphysics. In V. Munz, K. Puhl, & J. Wang (Eds.), *Language and the world, part one: Essays on the philosophy of Wittgenstein: Proceedings of the 32nd International Ludwig Wittgenstein Symposium in Kirchberg, 2009* (pp. 13–34). Frankfurt: Ontos Verlag.
Marion, M. (1998). *Wittgenstein, finitism, and the foundations of mathematics*. Oxford: Oxford University Press.
Moore, A. W. (2001). *The infinite*. London: Routledge.
Potter, M. (2000). *Reason's nearest Kin: Philosophies of arithmetic from Kant to Carnap*. Oxford: Oxford University Press.
Rodych, V. (2008). Mathematical sense: Wittgenstein's syntactical structuralism. In A. Pichler & H. Hrachovec (Eds.), *Wittgenstein and the philosophy of information* (pp. 81–103). Frankfurt: Ontos Verlag.
Rodych, V. (2011). Wittgenstein's philosophy of mathematics. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2011 Edition). URL http://plato.stanford.edu/archives/sum2011/entries/wittgenstein-mathematics/
Russell, B. (1919). *Introduction to mathematical philosophy*. London: Allen and Unwin.

Shapiro S. (2005). *The oxford handbook of philosophy of mathematics and logic*. Oxford: Oxford University Press.
Steiner, M. (1996). Wittgenstein: Mathematics, regularities, rules. In A. Morton & S. P. Stich (Eds.) *Benacerraf and His Critics* (pp. 190–212). Oxford: Blackwell
Steiner, M. (2009). Empirical regularities in Wittgenstein's philosophy of mathematics. *Philosophia Mathematica, 17*(1), 1–34.
Stern, D. (2004). *Wittgenstein's philosophical investigations. An introduction*. Cambridge: Cambridge University Press.
Urquhart, A. (2003). The theory of types. In N. Griffin (Ed.), *The Cambridge companion to Bertrand Russell* (pp. 286–309). Cambridge: Cambridge University Press.
Waismann F. (1979). *Wittgenstein and the Vienna circle: Conversations recorded by Friedrich Waismann*. In B. McGuinness (Ed., J. Schulte, B. McGuinness, Trans.). Oxford: Basil Blackwell.
Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford: Blackwell Publishing.
Wittgenstein, L. (1956/1991). *Remarks on the foundations of mathematics* (G.H. von Wright, R. Rhees, & G.E.M. Anscombe (Eds.), G. E. M. Anscombe, Trans.). Cambridge Massachusetts and London: MIT Press.
Wittgenstein, L. (1974). *Philosophical grammar* (R. Rhees (Ed.), A. Kenny, Trans.). Oxford: Basil Blackwell.
Wittgenstein, L. (1976/1975). *Wittgenstein's lectures on the foundations of mathematics*. Chicago: Chicago University Press.

Author Biography

Sorin Bangu is Professor of Philosophy at the University of Bergen, Norway. He works in contemporary philosophy of science and mathematics, and has interests in the history of analytic philosophy, focusing on Quine and Wittgenstein. He recently published a book, *The Applicability of Mathematics in Science: Indispensability and Ontology* (Palgrave Macmillan, 2012).

Wittgenstein's Color Exclusion and Johnson's Determinable

Sébastien Gandon

§1

The story has it that Wittgenstein, in 1928, discovered a flaw in the *Tractatus*, which led him to come back to philosophy and to revise his former system. What was the flaw? A very particular point of the Tractatusean doctrine, namely, the claim that elementary propositions are independent. At the time of the *Tractatus*, Wittgenstein held that elementary propositions could not contradict each other. Considering propositions about color exclusion, Wittgenstein stepped back in 1928, and claimed that elementary propositions like «a is red at t» and «a is blue at t» cannot both be true. This move has prompted a vivid discussion among scholars.[1] What exactly was Wittgenstein's point?[2] Did his criticism represent a new philosophical departure or should it be considered as a mere local adjustment which does not threaten the Tractatusean framework?[3] What were the reasons that explained

[1]The color exclusion issue was (much) discussed in the first scholarly works on Wittgenstein, from Ramsey (1923) to Pears (1987), including Anscombe (1959), Hacker (1972), Kenny (1972), Hintikka and Hintikka (1986). It seems that the issue has become a bit old fashioned since the nineties. See however Médina (2002), Moss (2012) for recent discussion.
[2]See for instance Médina (2002) for a critical analysis of this issue.
[3]Some commentators like Hacker (1972), Kenny (1972), Jacquette (1990) think that Wittgenstein (1929) represents a major break with the *Tractatus*; some like Hintikka and Hintikka (1986), Bertolet (1991) disagree; some others like Sievert (1989), Médina (2002 attempt to present an intermediate view.

S. Gandon (✉)
Blaise Pascal University, Clermont-Ferrand, France
e-mail: sgandon0@gmail.com; sgandon@orange.fr

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_10

Wittgenstein's change of mind?[4] And how to understand the importance conferred to what appears at first as a small change?

Here, I won't directly address these important and complicated issues. My aim is not to explain the evolution of Wittgenstein's thought during the years 1929–1930, but to draw attention to the conceptual similarities between Wittgenstein's (1929) discussion and Johnson's doctrine of determinable and determinate, expounded in the first volume of his *Logic* (1921). As the figure of William Ernest Johnson (1858–1931) is today forgotten, a brief presentation is not out of place.[5]

For nearly thirty years until his death, Johnson was Professor at Cambridge, where he taught philosophy, logic and probability. John Maynard Keynes and Frank Ramsey, whose works in probability theory bear witness to his influence, were among his students.[6] Johnson was also one of the few friends Wittgenstein had at Cambridge, both during his first stay in 1910–1913, and after his return.[7] Johnson's masterpiece is his *Logic* (in three volumes), which appeared in 1921–1924. With respect to the importance given it in the Aristotelian logic (see Sect. "§5" for more on this topic), the book was already dated at the time of its publication. But seeing Johnson only as a member of the British logic 'old guard', pushed aside by the *Principia Mathematica*, would be unfair and would not give credit to the richness of his thought. Indeed, many of Johnson's insights are today an integral part of philosophy.[8] This is especially the case with Johnson's doctrine of determinable and determinate (expounded in the Chap. 11 of the volume I), to which I will compare, in the sequel, Wittgenstein's (1929) discussion of color-statements.

My point in doing this comparison is not to uncover a hidden influence of Johnson on Wittgenstein. My perspective is here synchronic, not diachronic. Instead of considering 'Some Remarks of Logical Form' as a step in the journey from the *Tractatus* to the *Investigations*, I consider it as an integral part of a discussion, witnessed by Johnson (1921), which took place in Cambridge in the

[4]There are many proposals here. Let me just mention three important lines of interpretation. Some like Anscombe (1959), Médina (2002) consider that the point at stake in Wittgenstein (1929) is not the nature of color or degree statements, but the nature of logical possibility and linguistic rules. Some like Hintikka and Hintikka (1986), Sievert (1989), Bertolet (1991) link this debate to the nature of Tractatusean objects and to the role of Wittgenstein's phenomenology. As we will see, Wittgenstein (1921) makes a connection between the color exclusion issue and the impossibility for a body to be at different places at the same time—Ramsey (1923), Carruthers (1990), Moss (2012) try to develop this line.

[5]For biographical elements, see Broad (1931).

[6]In his *Treatrise on Probability*, Keynes refers to Johnson at several places (pp. 11, 68–70, 116, 124, 150–155). For Johnson's influence on Ramsey, see Sect. 7.

[7]See Monk (1990, 262), and below.

[8](Sanford 2011) notes that many of Johnson's linguistic innovations have entered the standard philosophical lexicon (determinate/determinable, occurrent/continuant, ostensive definition…). Johnson also made major contributions in probability theory and mathematical economics (for an assessment of Johnson's contributions to the theory of probability, see Zabell (1982); for an assessment of Johnson's contributions to mathematical economics, see Baumol and Goldfeld (1968).

twenties. This shift of emphasis does not, of course, render obsolete the genetic strategy; my bet is rather that it will complement the conclusions reached by more usual methods. Even if there were some good internal reasons which explained why Wittgenstein attached so much importance to color statement, there were also some contextual elements which, at that time in Cambridge, gave to this question a particular prominence.

I will first (Sects. §2–§4) summarize the discussions about the incompatibility of elementary propositions and about the color statements Wittgenstein made in the *Tractatus* and in 'Some Remarks of Logical Forms'. In a second part (Sects. §5–§7), I will speak about Johnson's doctrine of determinable and determinate, and compare it to Wittgenstein's views. In a third conclusive moment (Sect §8), focusing on an early work of Prior on the determinable and determinate, I will put Johnson's and Wittgenstein's discussions doctrine into a wider historical perspective.

§2

In the *Tractatus*, Wittgenstein held the view that elementary propositions are independent, i.e. it is possible for each elementary proposition to be true or false regardless of the truth or falsity of the others (4.211). Now, some statements, which seem to be no further reduced, are not independent. Consider the propositions *B*: '*a* is blue at t' and *R*: '*a* is red at t'. It is clear that *P* and *Q* cannot be true together; and yet, this incompatibility is prima facie not a truth-functional impossibility, since neither *B* nor *R* is a truth-function of more elementary propositions. Wittgenstein, in the *Tractatus*, is aware of this problem, since he explains (6.3751):

> For two colours, e.g., to be at one place in the visual field is impossible, and indeed logically impossible, for it is excluded by the logical structure of colour. (…) (It is clear that the logical product of two elementary propositions can neither be a tautology nor a contradiction. The assertion that a point in the visual field has two colours at the same time is a contradiction).

There are many divergent interpretations of this passage.[9] Two points are however non-controversial: the fact that the incompatibility of color statement is a logical, and not an empirical truth; and, the fact that Wittgenstein did not consider this incompatibility as a case running against his general thesis according to which elementary propositions are independent. How could Wittgenstein reconcile these two claims?

There is an easy way out: to renounce counting color statements as instances of elementary propositions. That *B* and *R* are apparently no further analyzable do not prove that they could not be analyzed as truth-functions of more fundamental

[9]See Allaire (1959) and Austin (1980) for a summary and a discussion of the first interpretations. See Médina (2002) and Sievert (1989) for a more up-to-date survey.

propositions. After all, as Wittgenstein said, 'language disguises the thought; so that from the external form of the clothes one cannot infer the form of the thought they clothe' (4.002). In 6.3751, Wittgenstein claims that color statements have a logical structure. He never claimed to have found this form. Wittgenstein, in the *Tractatus,* did not provide us with an effective analysis of color statements in which all incompatibilities would be reduced to truth functional propositions though he postulated that such a reduction could be done. Thus there is no incoherence between his general claim about elementary propositions and his idea that statements *B* and *R* exclude each other.

Things are, alas, more complicated. Wittgenstein went a step further in 6.3751, and, in so doing, spoiled the clarity of his thought. Immediately after having spoken about the 'logical structure of colour', he added:

> Let us consider how this contradiction presents itself in physics. Somewhat as follows: That a particle cannot at the same time have two velocities, that is, that at the same time it cannot be in two places, that is, that particles in different places at the same time cannot be identical.

As noted by Ramsey in his (1923) review of the *Tractatus*, the replacement of the phenomenological language (color statements) by a physical terminology (statements about places and velocities of bodies) does not help to reduce the incompatibility between *B* and *R* to a truth-functional contradiction. The fact that one and the same particle cannot be in two places at the same time is indeed not a truth-functional contradiction. And Ramsey confessed he did not see how to analyze the color statements in a way which makes them appear as truth-functionally linked to each other.[10] From our perspective, the most important point is not however the detail of Wittgenstein's proposal, but the fact that he did sketch, in the *Tractatus*, an (apparently flawed) particular analysis of the logical structure of color. This suggests that, contrary to what we have just said, Wittgenstein did not sharply distinguish the mere possibility of a truth-functional analysis of color statements from any particular view on the ultimate structure of color statement. What is exactly the status of his 6.3751 theory of color? And conversely, does it make sense to assert that a truth-functional analysis of a certain class of statements is possible, without giving any example of it? These questions still divide the scholars.[11] Without developing them any further, let me now turn to (Wittgenstein 1929).

---

[10]As (Proops 2013) emphasized, Wittgenstein was well aware of this point. In a *Notebooks* entry from August 1916 he remarks that: 'The fact that a particle cannot be in two places at the same time does look *more like* a logical impossibility. If we ask why, for example, then straight away comes the thought: Well, we should call particles that were in two places different, and this in its turn all seems to follow from the structure of space and particles.'

[11]Hintikka and Hintikka (1986) maintain that Wittgenstein had a clear idea of the ultimate structure of the world: sense data are *Tractatus*' simple objects. On the contrary, Médina (2002) and Proops (2013) maintain that Wittgenstein claimed that a reduction is possible, without knowing precisely how to do it.

§3

In 'Some Remarks on Logical Form', Wittgenstein explicitly rejects the position held in 6.3751. He now claims that some incompatibilities between propositions cannot, after all, be reduced to truth functional incompatibilities (1929, 168):

> I maintain that the statement which attributes a degree to a quality cannot further be analyzed, and, moreover, that the relation of difference of degree is an internal relation and that it is therefore represented by an internal relation between the statements which attribute the different degrees. […] The mutual exclusion of unanalyzable statements of degree contradicts an opinion which was published by me several years ago and which necessitated that atomic propositions could not exclude one another. I here deliberately say 'exclude' and not 'contradict', for there is a difference between these two notions, and atomic propositions, although they cannot contradict, may exclude one another.

Wittgenstein speaks here of a degree statement (as we will see below, this remark will turn out to have a great importance). But color statements fall under this umbrella. Thus (1929, 168):

> Take, for instance, a proposition which asserts the existence of a color R at a certain time T in a certain place P of our visual field. I will write this proposition 'RPT', and abstract for the moment from any consideration of how such a statement is to be further analyzed. 'BPT', then, says that the color B is in the place P at the time T, and it will be clear to most of us here, and to all of us in ordinary life, that 'RPT & BPT' is some sort of contradiction (and not merely a false proposition). Now if statements of degree were analyzable—as I used to think—we could explain this contradiction by saying that the color R contains all degrees of R and none of B and that the colour B contains all degrees of B and none of R. But … no analysis can eliminate statements of degree.

Why did Wittgenstein change his mind? Why did he hold in (1929), against 6.32751, that 'statements of degree' are not truth-functionally analyzable?

Once again, Wittgenstein seemed to be torn between two argumentative strategies. The first one has the following pattern: a particular truth-functional analysis of a degree statement, presented as a natural[12] hypothesis, is expounded, and then criticized in pages 167–168; from this, Wittgenstein concludes that 'no analysis can eliminate statements of degree.' As such, the reasoning is patently fallacious: that a particular analysis fails does not mean that no possible analysis exists. Happily, there is another argumentative line in Wittgenstein (1929). But before turning to this one, I would like to examine in more detail the natural hypothesis Wittgenstein sketches and finally rejects in 1929.

[12]Wittgenstein (1929, 167): 'One might think—and I thought so not long ago—that a statement expressing the degree of a quality could be analyzed into a logical product of single statements of quantity and a completing supplementary statement.'

Its basic claim is that one could not reduce an attribution of degree to a conjunction of statements attributing unit degree.[13] The 'natural' hypothesis is thus that a statement that a given entity E has two units of brightness b—'E(2b)'—could be equated to the conjunction of two statements attributing to E the degree b—'E(b) & E(b)'. This attempt fails because the truth-functional conjunction of the same proposition E(b) is logically equivalent to E(b). To capture the idea of addition of degree, one needs then to introduce a distinction between the first and the second occurrence of the unit degree 'b'. But this compels us to assume 'two different units of brightness', which, said Wittgenstein, 'is obviously absurd'.

Wittgenstein's argument has a venerable ancestor. Essentially the same reasoning had been used in the medieval dispute about the intensification and remission of form. Let me give an oversimplified summary of this debate.[14] In *Categoriae* (8, 10b 27–29), Aristotle claimed that some qualities admit the more and the less (i.e., degree). For instance, a white color, as a quality, can be more or less bright, and a white table can have more or less brightness. As Aristotle noted, admitting degree is not trivial—it is a property that does not belong to every category of being (for instance, it does not belong to substance), nor to every quality (it does not apply to geometrical figures). More importantly, a change in degree (an alteration of a quality) must be sharply distinguished from a quantitative change: an alteration is not the product of an addition (or diminution) of parts. In orthodox Aristotelian doctrine, a quality as such[15] has no part and could not be divided.

Now, for reasons coming from different sides,[16] some medieval thinkers from the thirteenth century came to challenge the Aristotelian distinction between alteration and addition. Let me quote Solère (2000, 585):

> For [the partisans of the additive theory], every quality whose intensity is increased acquires something new that it did not possess before. A distinct reality is added to the quality's preexisting degree, thereby creating a new unity …. Their view was developed by the so-called Oxford Calculators in the fourteenth century, and finally triumphed everywhere, and in particular, as we shall see, among the Jesuits of the sixteenth and seventeenth centuries.

[13]Wittgenstein (1929, 167): 'For let us call the unit of, say, brightness b and let E(b) be the statement that the entity E possesses this brightness, then the proposition E(2b), which says that E has two degrees of brightness, should be analyzable into the logical product E(b) & E(b). This analysis doesn't work, since E(b) & E(b) is equal to E(b), and not to E(2b); if, on the other hand, we try to distinguish between the units and consequently write E(2b) = E(b′) & E(b″), we assume two different units of brightness; and then, if an entity possesses one unit, the question could arise, which of the two—b′ or b″—it is; which is obviously absurd.'

[14]For a clear and informative account of the issue, I refer the interested reader to Solère (2000, 2001).

[15]Aristotle said that a quality can only be divided 'by accident', like for instance when a white surface is divided in different pieces. This division is a division of a quantity (the colored area), and not the division of the quality itself.

[16](Solère 2000, 2001) insists on the theological motivations.

The main opposition to this (anti-Aristotelian) view came from Thomas Aquinas. Solère summarizes Aquinas' objection in these terms (2000, 585–586):

> Thomas says that he cannot understand how charity (the reasoning would be the same for any other quality) could be augmented 'by addition of charity to charity.' For in the operation of addition, a distinct thing is added to another, and a distinction is either specific or numerical. But two cases of charity do not differ in their essence, and numerical difference depends exclusively on the diversity of the subjects in which accidents inhere.

This is in substance Wittgenstein's argument. Aquinas is even more precise, since he distinguishes between specific and numerical difference. Now, as the first quote above shows, this sort of criticism did not discourage the partisans of the additive theory, which finally triumphed. Indeed, several responses to the Thomist challenge emerged, which underlie many deep and complicated issues in Early modern philosophy.[17]

From a historical point of view at least, one can then recognize that the 'natural hypothesis' sketched by Wittgenstein (1929) played a central role in the Western philosophical tradition. But, far from destroying the additive theory, Aquinas' objection received different answers, that Wittgenstein did not even take the time to consider. This detour by the exotic medieval discussion about the 'latitude of form' reveals the weaknesses of Wittgenstein's analysis. It shows not only that the issue about the nature of the degree statement has a long history, but also that the intricacies of the problems raised are completely overlooked in Wittgenstein (1929). On the top of this, the general objection remains: that a particular truth-functional analysis of color statements fails does not mean that no such analysis can be found.

There is however a second, and much more plausible, interpretation of Wittgenstein's general argumentative strategy. In the *Tractatus*, Wittgenstein maintained that there is a truth-functional analysis of color statements, without expounding the logical structure of these statements. It seems that Wittgenstein became more and more skeptical about this idea. I. Proops, in his (2013), quotes a telling extract of the note taken by Moore (Wittgenstein is speaking here of generality, but his claim could be applied to color exclusion as well):

> There was a deeper mistake [in the *Tractatus*]—confusing logical analysis with chemical analysis. I thought '$(\exists x) \cdot fx$' *is* a definite logical sum, only I can't at the moment tell you which.[18]

Wittgenstein targets here the realist claim: according to the *Tractatus*, there is an ultimate structure of color statement as there is an ultimate structure of a chemical molecule. This logical realism grounded the claim that a truth-functional analysis could exist even if one does not provide the reader with an effective reduction. But as soon as the realist stance is given up, the idea that one can claim that a

[17]On this, see Solère (2001).

[18]Moore Archive, November 25, 1932, Entry 39 (quoted in Proops 2013). See also Moore (1955, 1–2).

truth-functional analysis is possible without actually giving it is called into question. Let me quote I. Proops who perfectly summarizes the idea:

> [In the *Tractatus*] Wittgenstein had supposed that there was a fact of the matter—unknown, but in principle knowable—about which logical sum '$(\exists x) \cdot fx$' is equivalent to. But because he had failed to specify the analytical procedure in full detail, and because he had not adequately explained what analysis is supposed to preserve, this idea was unwarranted. Indeed, it exemplified an attitude he was later to characterize as amounting to a kind of unacceptable 'dogmatism'.

This interpretation seems more reasonable than the one based on the criticism of the analysis of degree statement as truth-functional conjunction.

§4

That Wittgenstein changed his mind about color exclusion between 1921 and 1929 is undeniable. What his reasons were for doing so is however a question that still divides the commentators. Before explaining Johnson's doctrine of determinable and comparing it to Wittgenstein's position, I would like to highlight two features of Wittgenstein's evolution.

The first is quite basic, but, seemingly, has not been emphasized in the literature. In the *Tractatus*, the issue of color statements is never explicitly connected to the question of degree attributions. It is only in 1929 that Wittgenstein relates the two problems, without further explaining why the former should be considered as a *species* of the latter. That a shade of color is nothing other than a degree is not a trivial claim at all. Nothing shows that the author of the *Tractatus* ever endorsed this thesis. In 1921, the only occurrences of the concept of degree (Grad) are related to probability.[19] The idea that color statements are a *species* of a *genus* that I have called 'degree statements' should be considered as a new insight. In 'Some Remarks of Logical Form', Wittgenstein used the intricate phrase 'representation in which numbers (rational and irrational) must enter' to speak of the generic notion.[20] In *Philosophical Remarks*, Wittgenstein combined analyses of color statements (§§ 76–80, 86) with discussion of measurement (§§ 82–84). In both cases, the relevant features of color propositions are also attributed to a larger class of statements (the degree statements), and consequently, the analysis of color is always framed in the

[19]In *Tractatus* 4.464 and 5.155, Wittgenstein speaks of a 'Grad of Warscheinlichkeit'.

[20]Wittgenstein (1929, 165): 'We meet with the forms of space and time with the whole manifold of spatial (sic) and temporal objects, as colours, sounds, etc., etc., with their gradations, continuous transitions, and combinations in various proportions, all of which we cannot seize by our ordinary means of expression. And here I wish to make my first definite remark on the logical analysis of actual phenomena: it is this, that for their representation numbers (rational and irrational) must enter into the structure of the atomic propositions themselves'.

broader context of the analysis of degree. This is certainly a new ingredient of Wittgenstein's (1929) conception, and one which requires an explanation.

The second feature I would like to stress is more well-known: it concerns the connection between color exclusion and internal relation. Color relations are used by Wittgenstein in 1921 as in 1929 for exemplifying the notion of internal property.[21] There are however two different ways of conceiving this exemplification.

One way of explaining the connection is to focus on the relation between the general concept of color and the particular colors (or between one *species* of color, for instance, the red color, and its sub-*species*, the different shades of red). Let C be the concept of color, and let c_1, …, c_n all the different individual colors, the blue, the red, the yellow, etc. Wittgenstein, in the *Tractatus*, claims that the relation between C and c_1, …, c_n is internal, in the sense that one cannot empirically discover, as it were, a new color, and that, conversely, one will be immediately aware, placed in front of an incomplete color sampling, that one color is missing, even if one never saw a body of that color before.[22] The range of the possible particularizations of the concept of color (respectively, the range of the possible particularizations of a particular color concept) is fixed once and for all, independently of any empirical fact.

In 1929, Wittgenstein still adheres to this claim. But he extends this thesis by saying that attribution of color (or more generally, of degree) exhibits a certain form of completion (1929, 169):

> How, then, does the mutual exclusion of RPT and BPT operate? I believe it consists in the fact that RPT as well as BPT are in a certain sense *complete*. That which corresponds in reality to the function '()PT' leaves room only for one entity—in the same sense, in fact, in which we say that there is room for one person only in a chair. Our symbolism, which allows us to form the sign of the logical product of 'RPT' and 'BPT', gives here no correct picture of reality.

Resuming the notation introduced above, Wittgenstein's claim that color attribution is complete amounts to saying that only one of the particular colors c_1, …, c_n can 'saturate' the function '()PT'. Contrary to what he believed in the *Tractatus*, Wittgenstein considers now that color exclusion is primary data that cannot be reduced to truth-functions. Color concepts exhibit a particular logical feature

[21]Wittgenstein (1921, 4.123): 'A property is internal if it is unthinkable that its object should not possess it. (This shade of blue and that one stand, eo ipso, in the internal relation of lighter to darker. It is unthinkable that *these* two objects should not stand in this relation.)'.

[22]See Hume's wellknown passage about the missing shade of blue (*An Enquiry Concerning Human Understanding*, II) : 'Suppose therefore a person to have enjoyed his sight for thirty years, and to have become perfectly well acquainted with colours of all kinds, excepting one particular shade of blue, for instance, which it never has been his fortune to meet with. Let all the different shades of that colour, except that single one, be placed before him, descending gradually from the deepest to the lightest; it is plain, that he will perceive a blank, where that shade is wanting, said will be sensible, that there is a greater distance in that place betwixt the contiguous colours, than in any other. Now I ask, whether it is possible for him, from his own imagination, to supply this deficiency, and raise up to himself the idea of that particular shade, though it had never been conveyed to him by his senses? I believe there are few but will be of opinion that he can.'

(completion from which derives incompatibility) that is not captured in Russell's and Frege's notation. This means that color attributions must be endowed with their own specific logical forms, and not considered as ordinary predicate statements.

One should then carefully distinguish two claims about the internal properties of color- (and degree-) concept:

(A) the claim according to which a color-concept determines the range of its possible specifications
(B) the claim according to which the various determinations of color attributions exclude each other, and that color attributions cannot be reduced to predicate logic.

Claim (A) is endorsed both in 1921 and 1929, while claim (B) is adhered to only in 1929. It is important to distinguish between (A) and (B) because Johnson, to whom I will now turn, also made the distinction.

§5

Not much exists on Johnson's *Logic* (1921–1924) in the literature.[23] As I am only interested in one aspect of Johnson's thought, namely, the theory of determinable versus determinate, I won't deal at length with the whole project. But a few words are necessary to place the doctrine of determinable in its proper context.

Johnson's research must be seen as an investigation in the metaphysics of science rather than in logic. *Logic* II (published in 1922) is entirely devoted to an analysis of quantitative statement and causality, while *Logic* III (published in 1924) is concerned with the foundation of empirical science (induction and the distinction between body and mind).[24] *Logic* I (1921), which deals with logic in a restricted sense (with proposition, assertive force, generality, existential proposition, identity, etc.), appears then as a stepping stone to the epistemological and metaphysical considerations of the subsequent volumes. In the first Book, which I will focus on here, Johnson wanted first and foremost to present the tools that he used in the two other parts.

Book I is divided into fourteen chapters, each of which deals with a particular issue. There are obviously some links between them, but nowhere did Johnson explain and justify his progression. The reader, jumping from one chapter to the other, soon gets the impression that the work is not self contained, that its agenda comes from outside. This lack of explicit coherence can be partially explained by what I have just said: *Logic* I is just prolegomena for what comes after. But another

[23]Passmore (1968), Smokler (1967), Poli (2004) provide a good summary and an informed discussion of Johnson (1921–1924). Sanford (2011) and Prior (1949) discuss Johnson's theory of determinable.

[24]A Book IV, announced at the end of the introduction but which never came out, was supposed to give a subjective account of probability.

reason explains the piecemeal character of the book. Johnson was working from within the XIXth Century tradition of philosophical logic. His main references are (above all) the empiricist Mill, and the idealists Bradley and Bosanquet.[25] Even if these writers greatly diverged from each other, they shared a vast amount of presuppositions, among which the main ones were the centrality of Aristotelian syllogistic and the importance of Boolean algebraic logic. No surprise then if today readers have trouble getting their bearings: Johnson's milestones (Mill's, Bradley's and Bosanquet's thoughts) have completely disappeared from the contemporary philosophical curriculum.

There is however one important thread, which runs across the whole volume: Johnson's opposition to Bradley's monism. In his (1893), Bradley argued that any judgment distorts the structure of reality. In Bradley's language, a judgment is the union of a 'what' (which corresponds to the denotation of the subject) and a 'that' (which corresponds to a predicate). Now, Bradley maintained that no component of the judgment stands independently from the other in reality. The 'what', when isolated from the 'that', is 'a mere nothingness'; conversely, the predicate points toward something (the 'what') external to itself. For Bradley, it is only the 'abstractive character' of thought which is responsible for the segregation of the whole of experience in various 'what' and various 'that'.[26] In a sense, this is exactly what is expressed in a judgment, since a judgment is nothing other than an attempt to reunify what has been disconnected. But this attempt is bound to fail because the very act of judgment implies that the terms related are external and independent from each other, whereas their division is only the product of a prior thought activity.

Chapter II of Johnson (1921) is a criticism of this argument. For Johnson, the separation between the subject and the predicate is not a product of the activity of the thought. In perception, things are presented separately: experience has an immediate spatio-temporal structure, and space and time allow us to slice reality in different independent parts (1921, 22):

> [Mr Bradley's dictum 'distinction implies difference'] is exactly wrong: the assertion of 'otherness' does not presuppose or require a previous assertion of any relation of agreement or of difference. … The first important relation which will be elicited from otherness is, in fact, not any relation of agreement or difference at all, but a temporal or spatial relation; and thus the primitive assertion of otherness is only occasioned and rendered possible from the fact of separateness in presentation.

For Johnson, 'separateness is before relating' (1921, 21), and the structure of judgment does correspond to the structure of reality. Johnson maintains thus that a proposition can always be divided into a 'substantive' and an 'adjective', and that

[25]The central character of these references is clearly drawn in the introduction of (1921), especially in §8.

[26]See for instance Bradley (1893, 143–145). For a clear account of Bradley's logic, see Allard (2005).

this decomposition mirrors the structure of the fact.[27] In keeping with Aristotle, Johnson actually distinguished two kinds of substantive: the 'quasi-substantive' and the 'substantive proper', which 'seems to coincide with the category "existent"' (1921, xxxiv; see also 1922, xi–xiv). The idea is that, even if an adjective can be used as subject in a proposition, it will remain an adjective, since only limited spatio-temporal parts of experience can be considered as genuine substantives.[28] Johnson's defense of the strict Aristotelian orthodoxy does not come from a blind obedience to the tradition, but from the desire to defend rationalism against the threat represented by Bradley's idealist skepticism.[29]

Now, this distinction between 'substantive' and 'adjective' is important for us since the categories of determinable and determinate only apply to adjective (1921, xxxv):

> Adjectives are fundamentally distinguishable into determinable and determinates, the relation between which is primarily a matter of degree, a determinable being the extreme of indeterminateness under which adjectives of different degrees of determinateness are subsumed.

Owing to this limitation, Johnson often speaks about 'adjectival determination'. To illustrate what Johnson has in mind here, let's take the familiar example of the adjective color. Color is a determinable, under which fall different adjectives (the particular colors), which are the various determinates associated to this determinable. But a particular color (red for instance) is also a determinable *vis-à-vis* the different shades of colors that fall under it: the adjective carmine red is an adjectival determination of redness; and a shade (like carmine red) can also be considered as a determinable, if some further color-determinations are possible.

The basic idea is then simple: the distinction between determinable and determinate introduces a tree-structure within the province of adjectives. At the top of the tree, one finds the super-determinable (as for instance, color), that is, the adjectives that are not subsumable under any higher determinable; at the bottom, one finds the 'absolute determinate', the 'literal *infima species* under which no other determinate is subsumable' (1921, xxxv); in between, one finds the adjectives that are more determinate than some, but less determinate than others (as for instance the particular color red relatively to the adjective color on the one hand, and to the various shades of red on the other).[30] This tree-structure resembles then the scholastic Porphyry's tree, which represents a *species* by a *genus* and a *differentia*, and in which the process of differentiation continues until the lowest *species* (the *infima species*) is reached.

---

[27]It is to be noted that Johnson, in this framework, makes room for relations that he defines as special cases of adjectives and that he calls 'transitive adjectives'. See (1921, Chap. 13).

[28]On Johnson's theory of substantive and adjective, see Ramsey (1990, 9–10).

[29]Johnson was quite explicit on his wish to defend rationalism. He chose Aristotle's definition 'Man is a rational animal' as a motto of his *Logic*.

[30]On the relative character of the determinable/determinate distinction, see (1921, 177–178).

§6

The purpose of Chap. 11 is precisely to indicate in which respect the relation between determinable and determinate is distinct from the relation between *genus* and *species*—or what amounts to the same in which respect adjectival determination is distinct from class-membership (1921, 173–174):

> Superficially [the relation of a determinate to a determinable] appears to be the same as that of a single object to some class of which it is a member: thus two such propositions as 'Red is a colour' and 'Plato is a man' appear to be identical in form; in both, the subject appears as definite and singular, and, in both, the notion of a class to which these singular subjects are referred appears to be involved. Our immediate purpose is to admit the analogy, but to emphasise the differences between these two kinds of propositions, in which common logic would have said we refer a certain object to a class.

The theory of class is analyzed in Chap. 8. Two features are important: first, substantives but also adjectives can be elements of a class[31]; second, any class is determined by a property that every member of a class satisfies (that is, Johnson espoused an intensional view of class).[32] Thus, if Plato, Aristotle, Socrates are all considered as elements of the class of men, it is because they all share the same common property: that of 'being a man'.[33]

Now, the first difference Johnson makes between determination and class-membership consists precisely in the fact that determinates belonging to a same determinable do not share any common property. Red, yellow and green do not have a common property, a common point of agreement, which explains why we gather them under the common umbrella 'color'. As Johnson explained (1921, 175):

> If it is asked why a number of different individuals are said to belong to the same class, the answer is that all these different individuals are characterised by some the same adjective or combination of adjectives. But can the same reason be given for grouping red, yellow and green (say) in one class under the name color? What is most prominently notable about red, green and yellow is that they are different, and even, as we may say, opponent to one another. … In fact, the several colors are put into the same group and given the same name color, not on the ground of any partial agreement, but on the ground of the special kind of difference which distinguishes one color from another; whereas no such difference exists between a color and a shape.

---

[31]It would then be a mistake to align the distinction between adjectival determination and class membership applies to the distinction between adjective and substantive.

[32]In (1921, 122–124), Johnson makes a distinction between enumeration and class. An enumeration is purely extensional, contrary to a class, which is always determined via an adjective.

[33]It could be argued that adjectival determination resembles more set-inclusion than set-membership. After all, it would be more natural to translate 'Red is a color' as 'R $\subset$ C' (or as 'for all x, Rx implies Cx') than as 'R $\in$ C' (or as 'the entity R is C'). Johnson clearly makes the distinction between set membership ('$\in$') and set inclusion ('$\subset$') in (1921, 116–118), and one could regret that Johnson did not use the distinction in Chap. 11. But this lack does not undermine his analysis. As we will soon see, the way Johnson differentiates adjectival determination from class membership applies to class inclusion as well.

According to Johnson, the ground for grouping determinates under one and the same determinable is not any partial agreement between them, but 'the unique and peculiar kind of difference that subsists between the several determinates under the same determinable, and which does not subsist between any one of them and an adjective under some other determinable.' In other words, 'color' does not denote a particular property that each particular color has. 'Color' designates a system of ordered differences. The difference between red and yellow is not of the same kind than the difference between red and a trumpet blast, and this is the reason why the first two adjectives, but not the last one, fall under the determinable color.[34]

This first distinction between adjectival determination and class-membership is supported by a more fundamental one: in §2 of Chap. 11, determinates are said to 'emanate' from their common determinable.[35] To use Wittgensteinian terminology, one could say that the relation between determinable and determinate is 'internal': one cannot grasp the meaning of 'color', and then discover afterward that red is a color—while one can grasp the meaning of a class concept (like 'the apostles of Jesus') without knowing that an element (for instance, 'James') is an element of the class (1921, 177–178):

> Now no … class-name generates its members in [the way determinable generates its determinates]; take, for instance, 'the apostles of Jesus', the understanding of this class-name carries with it the notion 'men summoned by Jesus to follow him,' but it does not generate 'Peter and John and James and Matthew etc.,' and this fact constitutes one important difference between the relation of sub-determinate to super-determinate adjectives and that of general to singular substantives.

Bosanquet's idealistic theory of the concrete universal is probably at the source of this idea. As Foster (1931) explains, Bosanquet distinguished between the generic concept as differentiated in its *species* and the universal quality as predicable of its instances: 'the generic concept [unlike the universal quality], possesses a power of determining [or generating] the specific character in which it is realized' (1931, 9).[36] It seems thus that Johnson, when insisting on the internal character of

[34]On this example, see Johnson (1921, 190–191).

[35]See (1921, 174): 'Colour is not adequately described as undeterminate, since it is, metaphorically speaking, that from which the specific determinates, red, yellow, green, etc…, *emanate*.'

[36]See also (1931, 1): 'It is necessary from the beginning (…) to distinguish two different ways (or perhaps degrees) in which the universal may be conceived as actively determinant. First, the universal or 'Form' may be conceived not merely as a single universal characteristic which exists indifferently alongside others in a particular embodiment, but as a generic concept, which is active in determining (in 'generating') its own specific differentiations; in determining that is to say, which characters shall exist together to constitute an actual 'kind'. This view of the universal is expressed in the logical doctrine that the *differentia* of a *species* is not another character added to the generic character, but a differentiation of the generic character. It is typical of Aristotle, but not peculiar to him. It depends upon the insight, which Plato very clearly possessed, that the 'realm of forms' is not a mere multiplicity, but an intelligible system.'.

the relation between determinable and determinates, incorporated, in a non-idealistic setting, a key feature of Bosanquet's idealistic doctrine.

A third and last characterization of adjectival determination is developed in §3 of Chap. 11, in relation with the old scholastic rule according to which a decrease in extension is always accompanied by an increase in intension (and vice versa). This rule applies to *genus* and *species*: one should add to 'animal' the *differentia* 'rational' (i.e., increase the intension) in order to pass from the class of animal to the subclass of man (i.e., decrease the extension). Now, Johnson claimed that this rule does not hold in the case of adjectival determination (1921, 178):

> The phrase 'increase of intension' conjures up the notion of adding on one attribute after another, by the logical process called conjunction; so that, taking p, q, r, to be three adjectives, increase in intension would be illustrated by regarding p, q, r conjoined as giving a greater intension than p, q … We have now to point out that the increased determination of adjectival predication which leads to a narrowing of extension may consist —not in a process of conjunction of separate adjectives—but in the process of passing from a comparatively indeterminate adjective to a comparatively more determinate adjective under the same determinable. Thus there is a genuine difference between that process of increased determination which conjunctively introduces foreign adjectives, and that other process by which without increasing, so to speak, the number of adjectives, we define them more determinately.

To differentiate men from the other animals, one should refer to an external term (for instance the notion of rational). Applying this standard schema, one can wonder what differentiates red from other colors? And the answer seems to be: only redness itself. In the case of determinable and determinate, no addition of separate concepts (no *differentia*) can explain the decrease in extension.

Let me summarize Johnson's discussion. Despite their formal resemblance, 'Red is a color' and 'Plato is a man' do not have then the same logical form. The latter is an adjectival predication, said Johnson, while the former is an adjectival determination. Three arguments allow him to make this distinction: first, determinates falling under the same determinable do not share any common property; second, a determinable internally generates its determinates; thirdly, adjectival determination does not 'decrease extension' by 'increasing intension'.[37] Among these three arguments, the second is the most fundamental. It explains the third: if the determinable itself generates its determinate, no external addition is responsible for the process of determination. But it also grounds the first: in so far as it generates the

[37]Note that, as we have said above in footnote 34 this contrast applies equally well between adjectival determination and set-membership as between adjectival determination and set-inclusion.

totality of its determinates, the determinable is more a principle of differentiation than a common property.

Johnson's analysis does not stop here. Once the contrast between class-membership and adjectival determination is established, Johnson adds in §4 of Chap. 11 a crucial remark concerning the relations between 'adjectives under the same determinable' (1921, 181):

> If any determinate adjective characterises a given substantive, then it is impossible that any other determinate under the same determinable should characterise the same substantive: e.g. the proposition that 'this surface is red' is incompatible with the proposition 'this (same) surface is blue'.

We find here the same color exclusion issue from which we began. Two determinates belonging to the same determinable cannot be attributed to the same (genuine) substantive, i.e. to the same piece of spatio-temporal reality. For Johnson, this is the general principle underlying the color exclusion phenomena.

Strangely enough, Johnson does not provide us with any explanation of this fact. His reasoning seems to be this. Let's suppose for sake of convenience that the determinable C has only two determinates c_1, c_2. The adjective c_1 is not here a further property that an entity x, that has the property C could have; for x, to be c_1 is only a way of being C. And similarly, to be c_2, is only another way of being C. In other words, the incompatibility for an entity to be both at the same time c_1 and c_2 results from the fact that c_1 and c_2 are not two independent properties, but two determinations of the same property C. Once the determination has taken place, it is no longer possible to determine the determinable C in any other way, without destroying the previous determination.

In any event, this incompatibility between determinates (of the same determinable) plays a fundamental role in Johnson's logical thought. For instance, in Chap. 5 of (1921), Johnson explains that each negation has a ground: in denying that an object x has a certain adjective f, one always attributes to it the generic adjective which is the determinable of f: 'When we deny of a flower that it is red, we are at least judging that it has some color' (1921, 67). Johnson goes very far in this direction, since he reduces truth-functional contradiction to the incompatibility between determinates. Let me quote (1921, 15):

> We may illustrate the relation of incompatibility amongst adjectives by *red* and *green* regarded as characterizing the same patch. It is upon this relation of incompatibility that the idea of *not-red* depends; for *not-red* means some adjective incompatible with *red*, and predicates indeterminately what is predicated determinately by *green*, or by *blue*, or by *yellow*, etc.

This completely reverses the picture presented in the *Tractatus*: far from being reducible to a truth-functional contradiction, incompatibility between determinates is the basis on which an account of negation is grounded.[38]

§7

Johnson's key distinction between adjectival determination and set-membership is close to Wittgenstein's idea that degree statements have their own specific logical form, which should be sharply distinguished from the predicative form. Moreover, in both cases, the phenomena of color (resp. determination) exclusion is taken as a sign that shows the heterogeneity between statement of degree (adjectival determination) and usual predication. With respect to their main thesis, there is thus a convergence of view between Wittgenstein (1929) and Johnson (1921): in both works, color (resp. determination) exclusion is treated as a primary datum, which cannot be reduced to Frege's and Russell's logic.

This agreement between Wittgenstein and Johnson extends even to the way they argue their main claim. Indeed, the distinction made in Sect. 5 above between the two claims (A) and (B) was also made by Johnson. In the last chapter of *Logic* I, Johnson presented a 'formal' characterization of adjective determination in the form of four laws relating substantive, determinate and determinable (1921, 237):

1. *Principle of Implication*: If s is p, where p is a comparatively determinate adjective, then there must be some determinable, say P, to which p belongs, such that s is P.
2. *Principle of Counterimplication*: If s is P, where P is a determinable, then s must be p, where p is an absolute determinate under P.
3. *Principle of Disjunction*: s cannot be both p and p', where p and p' are any two different absolute determinates under P.

[38]In the conclusion of Chap. 11, Johnson underlines two important applications of the notion of determinable. First, determinables play a key role in the theory of causality—thus for instance, the idea that an event is caused by another one according to some fixed laws is reformulated in these terms (1921, 243): 'taking any determinable P, the determinate value which it assumes in any manifestation is determined by the conjunction of a finite number of determinables A, B, C, D (say) such that any manifestation that has the determinate character *abcd* (say) will have the determinate character p (say)'. Second determinables play a central role in probability. Johnson uses determinable in the definition of the sample space of a random event—see (1921, 178–183): 'In the problem before us, we shall be concerned with a certain adjectival determinable P which has α determinate values—p_1, p_2, …, p_α—and shall proceed to consider M instances, each of which is characterised by one or other of these determinate characters. Any actual set of occurrences of length M will exhibit a certain proportion among the determinate characters;—m_1 occurrences of p_1, m_2 of p_2 … m_α of p_α (say), where $m_1 + m_2 + \cdots + m_\alpha = M$… *Combination-Postulate*: In a total of M instances, any proportion, say $m_1 : m_2 : \ldots : m_\alpha$, where $m_1 + m_2 + \cdots + m_\alpha = M$, is as likely as any other, prior to any knowledge of the occurrences in question.'

4. *Principle of Alternation*: *s* must be either not-*P*; or *p* or *p′* or *p″*. … continuing the alternants throughout the whole range of variation of which *P* is susceptible – *p*, *p′*, *p″* … being comparatively determinate adjectives under *P*.

Let's focus on the two last principles. In the *Tractatus*, Wittgenstein endorsed the principle of alternation, since he maintained that the range of variation of the concept of color is fixed once and for all, independently of any empirical fact. On the other hand, Wittgenstein rejected, in 1921, the principle of disjunction. Incompatibility between determinates, then, was not considered as a primitive logical fact. Thus, in Johnson (1921), we find both the presence and the distinction of the two claims about internal relation of color terms we spoke about before: Johnson's principle of alternation is a version of (Wittgenstein's 1921) (A) claim, while his principle of disjunction is a version of (Wittgenstein's 1929) (B) claim. In 'Some Remarks of Logical Form', the central issue is to know whether Johnson's principle of disjunction should be considered, alongside with (1), (2) and (4), as a primitive logical principle.

There is another important similarity between Wittgenstein and Johnson: they both consider color statements as a part of a larger whole, called degree statements in Wittgenstein (1929), and adjectival determination in Johnson (1921). One of Johnson's most important insights is certainly to have grouped a vast family of propositions exhibiting the same kind of logical behavior under a common umbrella. Nowadays of course, philosophers have no trouble with identifying this heterogeneous set of statements as a particular group, whose logical features are more or less captured by Johnson's characterization. But before Johnson (1921), nobody[39] thought to merge statements about colors, measurements, and finite numbers,[40] nor to contrast them as a whole with the class of predicative statements. Johnson should be credited for this achievement. And as I have said (see Section §4), Wittgenstein followed Johnson's path, since, in 1929, he considered for the first time color statements as a *species* of degree statements. This inclusion of color statements in a larger category did not go without saying in the twenties. Johnson had to argue at length to show that his new classification was relevant.

The conceptual resemblances between Wittgenstein and Johnson's views should not hide however that there are important differences between the two. As we saw, Johnson's thought is rooted in the tradition of XIXth Century algebraic logic, which Frege and Russell (and then Wittgenstein) broke with. Wittgenstein emphatically rejected the Aristotelian analysis of proposition in terms of subject and predicate, which constituted the basis of Johnson's construction. Now, concerning the doctrine of determinable, adjectival determination has a wider scope than Wittgenstein's degree statement. For instance, an arithmetic sentence like '2 is

[39]In Sect. 8, I will qualify this claim.

[40]Finite numbers are considered by Johnson as determinates of the determinable 'natural number'.

even' is considered by Johnson as having the same logical form as the statement 'red is a color', which is of course not the case in Wittgenstein.[41] Finally, Johnson and Wittgenstein do not attribute to adjectival determination the same logical status. Both consider that the relation between the determinate and determinable is internal; but Johnson seems to consider that 'red is a color' is a genuine proposition that is a priori true; for Wittgenstein, on the contrary, 'red is a color' is a grammatical rule, not a proposition. But this is a vexed issue both because Wittgenstein's views on the status of grammatical rules evolve during the period, and because Johnson developed a theory of 'structural propositions', which is close to Wittgenstein's Tractatusean 'senseless propositions'.[42]

These divergences, if real, are however marginal in regard to the central fact that no one at the time, except Wittgenstein and Johnson, did draw attention to the contrast between predication and adjectival determination (to resume Johnson's terminology). Given this similarity, the question raises itself: did Wittgenstein read Johnson?

Wittgenstein did not refer to Johnson in any of his published writings.[43] And if the picture Monk gives of their relation is correct, one can well imagine why (1990, 262):

> Wittgenstein maintained an affectionate friendship [with Johnson], despite the intellectual distance that existed between the two. Wittgenstein admired Johnson as a pianist more than as a logician, and would regularly attend his Sunday afternoon 'at homes' to listen to him play. For his part, though he liked and admired Wittgenstein, Johnson considered his return a 'disaster for Cambridge'. Wittgenstein was, as he said, 'a man who is quite incapable of carrying on a discussion'.

Wittgenstein would have appreciated Johnson the man and the pianist, but he would have despised Johnson the philosopher. I doubt, however, that this tableau is correct. In another passage, Monk tells us (1990, 115) that, before the war, Wittgenstein had arranged with Keynes to donate 200 £ a year to a research fund administrated by King's College to help Johnson continue in his work on logic. More importantly, Wittgenstein, having just received the finished copies of the *Tractatus*, wrote to Ogden, on 15 November 1921: 'I should like to know what [Johnson] thinks about it. If you see him please give him my love.' (1990, 213). Wittgenstein's behavior, as reported by Monk, does not to square with the idea that Wittgenstein completely disregarded Johnson's intellectual capacities.

Even if Wittgenstein did not read *Logic* I, it is likely that he heard about Johnson's theory of determinable from Ramsey, who visited him in Austria in 1923

[41] In the same vein, the Middle Wittgenstein would have refused to ground the truth-functional logic (and notably the analysis of negation) on the doctrine of determinable.

[42] See Johnson (1922, 13–17). For an analysis of Johnson's structural proposition, see Prior (1949, 178–183).

[43] Sect. 44 of Wittgenstein (1964), about formally certified propositions seems to be about the theory expounded in Johnson (1921, Chap. 4). I thank Mauro Engelmann to have directed my attention to this passage.

and 1924, and in Cambridge in 1929.[44] Ramsey held Johnson's philosophical works in high esteem. He carefully read Johnson (1921)[45] and wrote a very positive review of Johnson (1922).[46] Recall also that one of the objections Ramsey formulated in his review of the *Tractatus* concerned Wittgenstein's treatment of color incompatibility. The topic probably came out in the conversations between Wittgenstein and Ramsey and, owing to Ramsey's familiarity with Johnson (1921), it seems natural to believe that the theory of determinable was at least mentioned.

As a matter of fact, one passage of Wittgenstein's lectures (taken from the notes of John King and Desmond Lee) given in 1930 clearly alludes to Johnson's doctrine (Wittgenstein 1989, 13):

> Johnson says that the distinguishing characteristic of colours is their way of difference from each other. Red differs from green in a way that red does not differ from chalk. But how do you know this? "It is formally not experimentally verified" (W.E. Johnson, Logic I, 56). But this is nonsense. It is as if you were to say you could tell if a portrait was like the original by looking at the portrait alone. Language shows the possibility of constructing true and false propositions, but not the truth or falsehood of any particular prop. So there are no true a priori propositions (mathematical propositions so called are not propositions at all).

The remark is clearly critical: Wittgenstein refused to consider an adjective determination like 'red is a color' as an (a priori) proposition.[47] As I have just said, Johnson's view on this topic is not as simple as Wittgenstein seems to think. But the interesting point lies elsewhere: the two first sentences show that Wittgenstein knew Johnson's doctrine of determinable.

This being granted, I don't want to suggest that Wittgenstein, in 1929, slavishly followed Johnson, without having paid due credit to his work. As we have seen, the divergences between the two philosophers are important. And it is also possible that Wittgenstein came to his conclusions by himself. If the comparison between Wittgenstein (1929) and Johnson (1921) is worth making, it is for reasons which have nothing to do with a possible influence of the latter on the former. Let me explain this in the next section.

---

[44]On the latter, see Moore (1954a, b).

[45]There is, in the Ramsey's archives in Pittsburgh, a quite detailed (26 handwritten pages) summary of the first eleven chapters of Johnson (1921). See Box 5, Folder 22–25, http://digital.library.pitt.edu/u/ulsmanuscripts/pdf/31735044223216.pdf.

[46]See Ramsey (1922). Ramsay referred to Johnson several times in his work and papers. See Glavotti's comment on the relationship between Johnson and Ramsey in (1991, 20–21).

[47]See (1989, 12): '"Primary colour" and "colour" are pseudo-concepts. It is nonsense to say "Red is a colour," and to say "There are four primary colours" is the same as to say "There are red, blue, green, and yellow." The pseudo-concept (colour) draws a boundary *of* language, the concept proper (red) draws a boundary *in* language.'

§8

A striking feature of Wittgenstein's writings is their self-centered character. In the transition period, Wittgenstein opposed his new thoughts to his own Tractatusean views, but he made very few references to other philosophers. Of course, Wittgenstein discussed the works of Frege and Russell, but his relations to them could not be dissociated to his relation to his own Tractatusean past. This idiosyncratic feature raises a question: as a commentator, should we follow Wittgenstein and content ourselves to relate the new developments to the old ones? Or should we break with this attitude, and embed Wittgenstein's thoughts within a non-Tractatusean background?

In this paper, we took the second branch of the alternative. This poses, of course, a problem: as Wittgenstein did not himself refer to the philosopher he is compared to, how to justify the rapprochement? Superficial connections could be found between nearly any pair of philosophical works. Without the safety net of textual evidences, what is the value of such comparisons? To this genuine threat, there is, I think, no general defense. I can only repeat the reasons which led me think that this rapprochement between Johnson and Wittgenstein is justified: the topic of the comparison is well defined and limited in its scope; on the issue of color exclusion, there is an astonishing convergence between the two philosophers (see Sect §7); Johnson and Wittgenstein belonged to the same Cambridge world, knew each other well, and had a common philosophical sparring partner, namely F.P. Ramsey; finally, Wittgenstein refers to Johnson's adjectival determination in one of his 1930 courses.

I would like now to tell why this rapprochement is worth making. Contrary to Wittgenstein, Johnson was an academic philosopher, who considered his work as a continuation and refinement of what had been done before him. His theory of determinable, in particular, is not viewed as something completely new, but as a resumption of some doctrines which were held in the past. Relating Johnson (1921) to Wittgenstein (1929) allows us to articulate the somewhat idiosyncratic color exclusion issue in Wittgenstein with a long-term process in the history of occidental philosophy. Let me explain this point by focusing on Prior's paper 'Determinables, determinates, and determinants', published in 1949. Prior's official aim is to develop Johnson's insights. But, in the first two sections of his article, Prior set about locating the theory of determinable within the history of philosophical logic.

According to Prior, in the traditional Aristotelian and scholastic doctrine, when the name of any *species* is rightly defined, its meaning is analysed into two parts with clearly distinguishable logical functions. The *genus* is the 'material' or 'determinable' part of the meaning (*pars determinabilis essentiae*); it tells us the sort of thing that is meant. The *differentia* is the 'formal' or 'determining' part of the meaning (*pars determinans essentiae*); it tells us by what 'form' or quality this *species* is marked off from others of the same *genus*. According to Prior then, Aristotle's doctrine contains two ideas: first, that the *differentia* is an additional term which is added to the *genus*; second, that there is an asymmetry between the *genus*

and the *differentia*. Now, Prior's main claim is that these two ingredients do not fit together. One can either choose to maintain the distinction between the *differentia* and the *genus*, but one has then to renounce the asymmetry between the two. Or one can decide to maintain the asymmetry, but it would then be to the detriment of the distinction between *genus* and *differentia*. According to Prior, both paths have been followed since Aristotle.

Leibniz and the Boolean tradition have explored the first avenue. In this framework, one considers the definition of man as a conjunction of two terms, 'animal' and 'rational'. As conjunction is a symmetrical operator, the Aristotelian idea that there is an asymmetry between *genus* and *differentia* is given up. In *Nouveaux Essais* (III, iii, 10), Leibniz claims thus that 'very often the *genus* may be changed into the *difference* and the *difference* into a *genus*; [as] for example, the square is a regular quadrilateral, or rather a four-sided figure that is regular'. Prior also mentions Couturat, Schröder and Keynes, for which 'a determinant is any one of the elements combined in a logical conjunction' (1949, 3).

Now, according to Prior, Johnson presents us with a consistent elaboration of the second branch of the post-Aristotelian alternative (1949, 6):

> The point about determinable *genera* and determinate *species*, in Johnson's sense of 'determinable' and 'determinate', is that there is no distinct *differentia* employed in passing from the former to the latter.

One recognizes here the third characterization developed by Johnson in Chap. 11 of his (1921). Unlike what happens in Leibniz and the Boolean tradition, the *differentia* is not considered as an independent term. The *genus* is given first, and *differentia* is nothing other than a determination of the *genus*. Johnson is however not the first to have taken this path. As Prior explains (1949, 7):

> In a brief but suggestive footnote (*Logic* III. V. 1) Johnson invites comparison between his own terminology and that of Descartes and Spinoza: 'What I call a determinable is almost equivalent to what they call an attribute, and my determinate almost equivalent to their mode of an attribute.' If Leibniz brought consistency into the medieval view of the *distinctness* of *genus* and *differentia* by making the relation between them symmetrical, Spinoza brought consistency into the medieval view of the asymmetry of the relation by denying them their distinctness. Spinoza's 'modes' are not new qualities added to his 'attributes' … but determinations of them.

As Prior notes, after Spinoza and Descartes, the 'assymetry view' did not disappear from the scene: philosophers as different as Locke, Bradley and Bosanquet espoused this perspective. But, unlike its rival, this approach has never been articulated in one coherent and recognizable doctrine.[48] According to Prior, the

[48]Prior notes that 'a tradition which runs from Descartes and Spinoza through Locke to Bradley and Bosanquet is something of a philosophical curiosity, the appearance of Locke in such company being particularly strange' (1949, 7).

great merit of Johnson is having succeeded in crystallizing in one unified whole what was, before him, spread across different works.

For Prior then, the contrast drawn by Johnson between predication and adjectival determination has a long history: it is directly related to the ambiguity of the Aristotelian doctrine, and to the idea that the standard Leibnizian way of resolving the issue should be resisted. Even if one can dispute the historical accuracy of this description (was Aristotle's theory of *genus* and *differentia* really ambiguous? Were Spinoza and Leibniz really opposed?), this story, taken as an interpretation of Johnson (1921), is convincing, since Johnson himself made the connection between adjectival determination and the scholastic doctrine of the *fundamentum divisionis*, Descartes' and Spinoza's view of the attribute, and Bosanquet's theory. The distinction between adjectival determination and predication is not an isolated innovation, but something which is connected to issues which are central in the history of the Aristotelian logic.

Now, let me come back to Wittgenstein. The criticism of Frege's and Russell's logic is one of the main lines of *Philosophical Remarks*. The two philosophers are criticized for having reduced any kind of generality to the standard truth-functional one. Here is a representative sample of what can be found in Wittgenstein (1964, 119):

> One difficulty in the Fregean theory is the generality of the words 'concept' and 'object'. For even if you can count tables and tones and vibrations and thoughts, it is difficult to bracket them all together. Concept and object: but that is subject and predicate. And we have just said that there is not just one logical form which is *the* subject predicate form.

The developments on color and degree exclusion in Wittgenstein (1929) squares within this framework: color statements (to use Johnson's terminology) is just one among the various forms that should not be reduced to 'the' standard truth functional predication. And we touch here one limit of this interpretative grid. By embedding Wittgenstein's discussion within the context of a discussion of Frege's and Russell's logic, one understands the nature of his criticism, but one does not explain the importance Wittgenstein attached to the logic of color statements. From this perspective, the 'language game' of color and degree has nothing special: it is only one of the various pieces of language whose formal behavior is not truth-functional.

Our detour by Johnson and Prior allows us to complement this approach and to give a reason why color exclusion matters after all. Be it conscious or not, Wittgenstein, in contrasting degree attribution with predication, recapitulated a deep and vast movement, which crossed over the entire history of philosophical logic, and which opposed the partisans of a Leibnizian combinatorial approach to the upholders of a Spinozist asymmetrical view. Frege and Russell belonged to the former group, while Johnson and Wittgenstein (in 1929) belonged to the latter. In this broader context, the 'language game' of color statements played a central role, since it provided the partisans of the Spinozist approach with their main example.

Color and degree statements were crucial because they were cases where the difference between predication and adjectival determination were the most apparent.

Let me summarize my point. The main insight of Wittgenstein (1929), namely, the idea that degree statements have a distinctive logical form, is central in Johnson (1921). Prior (following Johnson) traced this insight back to Spinoza and Descartes; he also attributed it to Locke and Hume, and noted that it cropped up again in the works of Bradley and Bosanquet. Then, far from being idiosyncratic, Wittgenstein's argument in 1929 resumed a move made by many before him. To say this, is not undermining the strength of Wittgenstein's development. On the contrary, it is to connect Wittgenstein's thought to a major current in the history of philosophy, and to help locate his position within a broader conceptual space. As I have noted, comparing the works of two philosophers who do not refer to each other is definitely taking a risk. But making any comparison between philosophers conditional to the existence of cross-references is also a hazardous strategy. In the case at stake, refusing to relate Wittgenstein to Johnson would lead us to isolate Wittgenstein's thought.

The wish to de-mythologize Wittgenstein by connecting his work to the long term history of philosophy[49] has another benefit: it renders possible the use of Wittgenstein in contemporary philosophical discussion. Let me quote Kit Fine's introduction to his recent axiomatization of the determinable/determinate logic (2011, 161):

> In the *Tractatus*, Wittgenstein took the atomic propositions, by which the world is to be described, to be completely independent of one another. But he later revised his view (Wittgenstein 1929) and allowed that the atomic propositions might exhibit the kind of dependence that is characteristic of the way in which different determinants of a given determinable are exclusive of one another. Our question might therefore be put in the form: how in the most abstract terms should we conceive of the post-Tractarian world?

From the standard approach, which looks at Wittgenstein's middle period writings from a genetic perspective, Fine's presentation is unintelligible. Wittgenstein (1929) is not a elaboration of a theory of determinable. For a purist, Fine is here projecting his own preoccupations on Wittgenstein. It is to challenge this puritan attitude that this paper has been written. From a more encompassing historical perspective, Fine's presentation can be vindicated.

References

Allaire E. B. (1959). Tractatus 6.3751. *Analysis*, *19*(5), 100–105. Reprint in S. Shanker (1986) vol. 1, pp. 202–206.

Allard J. (2005). *The logical foundations of Bradley's metaphysics. Judgment, inference, and truth*. Cambridge University Press: Cambridge.

[49]In this, I follow the suggestion made by Marion in his (2004).

Anscombe E. (1959). *Introduction to Wittgenstein's tractatus*. UK: Hutchinson & Co Ltd.
Austin J. (1980). Wittgenstein's Solutions to the color exclusion problem. *Philosophy and Phenomenological Research*, *41*, 142–149. Reprint in S. Shanker (1986) vol. 1, pp. 207–212.
Baumol W. J., & Goldfeld S. D. (1968). Precursors in mathematical economics: An anthology. *London School of Economics*.
Bertolet, R. (1991). Elementary proposition, independance, and pictures. *Journal of Philosophical Research, 16*, 53–61.
Bradley F. H. (1883). *The principles of logic*. London: Oxford University Press. Second edition, revised, with commentary and terminal essays, London: Oxford University Press, 1922.
Bradley, F. H. (1893). *Appearance and reality*. Oxford: Clarendon Press.
Broad, C. D. (1931). William Ernest Johnson. *Proceedings of the British Academy, 17*, 491–514.
Carruthers, P. (1990). *The Metaphysics of the tractatus*. Cambridge: Cambridge University Press.
Fine, K. (2011). An abstract characterization of the determinate/determinable distinction. *Philosophical Perspectives, 25*(1), 161–187.
Foster M. B. (1931). The Concrete Universal: Cook Wilson and Bosanquet. *Mind*, *40*(157), 1–22.
Hacker, P. M. S. (1972). *Insight and illusion*. Oxford: Clarendon Press.
Hintikka, J., & Hintikka, M. (1986). *Investigating Wittgenstein*. Oxford: Blackwell.
Jacquette, D. (1990). Wittgenstein and the color incompatibility problem. *History of Philosophy Quarterly, 7*, 353–365.
Johnson W. E. (1921). *Logic, part I*. Mineola: Dover Publication. Reprint 1964.
Johnson W. E. (1922). *Logic, part II*. Mineola: Dover Publication. Reprint 1964.
Johnson W. E. (1924). *Logic, part III*. Mineola: Dover Publication Reprint 1964.
Kenny, A. (1972). *Wittgenstein*. Harmondsworth: Penguin.
Marion, M. (2004). *Ludwig Wittgenstein. Introduction au Tractatus logico-philosophicus*. Paris: Presses Universitaire de France.
Médina, J. (2002). *The unity of Wittgenstein's Philosophy: Necessity, intelligibility and normativity*. Albany: State University of New York Press.
Moore, G. E. (1954a). Wittgenstein's lectures in 1930–33. *Mind, 63*(249), 1–15.
Moore, G. E. (1954b). Wittgenstein's lectures in 1930–33. *Mind, 64*(251), 289–316.
Moore, G. E. (1955). Wittgenstein's lectures in 1930–33. *Mind, 64*(253), 1–27.
Monk R (1990). Ludwig Wittgenstein: The Duty of Genius, Peguin Books.
Moss S. (2012). Solving the Color Incompatibility Problem, *Journal of Philosophical Logic* vol. 41, no. 5 (2012): 841–51.
Passmore J. (1968). *A hundred years of philosophy* (2nd ed.). Harmondsworth: Penguin.
Pears D. (1987). The False Prison, vol. I, Oxford University Press.
Poli, R. (2004). W. E. Johnson's determinable-determinate opposition and his theory of abstraction. *Poznan Studies in the Philosophy of the Sciences and the Humanities, 82*(1), 163–196.
Prior A. N. (1949). Determinables, determinates, and determinants. Part I. *Mind*, *LVIII*, 1–20; Part II. *Mind*, *LVIII*, 178–194.
Proops I. (2013). Wittgenstein's logical atomism. In N. Z. Edward (Ed.), *The stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/sum2013/entries/wittgenstein-atomism/
Ramsey F. P. (1922). Review of W. E. Johnson's logic, part II. *The New Statesman 19*, 469–470,
Ramsey, F. P. (1923). Critical notice of L. Wittgenstein's Tractatus. *Mind, 32*(128), 465–478.
Ramsey F. P. (1990). In D. H. Mellor (Ed.), *Philosophical papers*. Cambridge: Cambridge University Press.
Ramsey F. P. (1991). In M. C. Galavotti (Ed.), *Notes on philosophy, probability and mathematics*. Bibliopolis.
Sanford D. H. (2011). Determinates vs. determinables. *Stanford Encyclopedia of Philosophy* (Online).
Sievert, D. (1989). Another look at Wittgenstein's color exclusion. *Synthese, 78*, 291–318.
Shanker S. (Ed.). (1986). *Ludwig Wittgenstein: Critical assessments* (Vols. 1–5). London: Croom Helm.

Smokler H. E. (1967). Johnson, William Ernest. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 4, pp. 292–293). UK/New York: McMillan/Free Press.
Solère J.-L. (2000). Plus ou moins: le vocabulaire de la latitude des formes, in *L'Elaboration du vocabulaire philosophique au Moyen Age*, éd. J. Hamesse et C. Steel eds., Turnhout, Brepols, "Rencontres de Philosophie médiévale" n°8, 437–488.
Solère J.-L. (2001). The question of intensive magnitudes according to some Jesuits in the sixteenth and seventeenth centuries, *The Monist*, *84*(4), 582–616.
Wittgenstein L. (1916). *Notebooks 1914–16* (G. E. M. Anscombe, Trans.). Oxford: Blackwell.
Wittgenstein L. (1921). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuinness, Trans., 1961). London: Routledge.
Wittgenstein, L. (1929). Some remarks on logical form. *Proceedings of the Aristotelian Society, 9*, 162–171.
Wittgenstein L. (1964). In R. Rhees (Ed.), *Philosophical remarks*. Oxford: Basil Blackwell. Reprint by R. Hargreaves & R. White Trans., 1975.
Wittgenstein L. (1989) *Wittgenstein's lectures, Cambridge, 1930–32. From the notes of John King and Desmond Lee*. Chicago: University of Chicago Press.
Zabell S. L. (1982). W. E. Johnson's 'sufficientness' postulate. *The Annals of Statistics*, *10*(4), 1090–1109.

Author Biography

Sébastien Gandon is professor at the Université Blaise Pascal in Clermont-Ferrand, France. He specializes in history of analytic philosophy, especially in the interactions between the scientifical background and the philosophical thoughts in Frege, Russell, Wittgenstein, Carnap, and also in other less-known figures such as Wiener and Johnson. He is the author of *Russell's Unknown Logicism*.

The Concept of "Essential" General Validity in Wittgenstein's *Tractatus*

Brice Halimi

Logical Generality and Logical Validity

This chapter is a comment on the following passage of the *Tractatus*:

> 6.1231. The mark of a logical proposition is *not* general validity.
>
> To be general means no more than to be accidentally valid for all things. An ungeneralized proposition can be tautological just as much as a generalized one.
>
> 6.1232. The general validity of logic might be called essential, in contrast with the accidental general validity of such propositions as 'All men are mortal'. Propositions like Russell's 'axiom of reducibility' are not logical propositions, and this explains our feeling that, even if they were true, their truth could only be the result of a fortunate accident.[1]

The above text puts forward a new characterization of the "general validity of logic," which implies a new characterization of both logical generality and logical validity. The Tractarian concept of essential general validity is not clear by itself. Few commentators, however, have dwelled upon its exact content. Anscombe does not analyze it. Robert Fogelin, in his chapter devoted to necessity in the *Tractatus*,[2] focuses on 6.37 ("The only necessity that exists is *logical* necessity") and does not explain the peculiar concept of logical validity to which Wittgenstein alludes. Michael Morris skims past the issue.[3] Peter Hacker mentions "the *essential* validity of logic," but only as a rebuttal of Russell, and without explaining its positive

[1] Wittgenstein (1974, p. 63).
[2] Fogelin (1976, Chap. vii, pp. 78–83).
[3] Morris (2008, pp. 237–240).

B. Halimi (✉)
Université Paris Ouest Nanterre La Défense (IREPH) & SPHERE, Paris, France
e-mail: bhalimi@u-paris10.fr

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_11

content apart from referring 6.1231 back to 6.113.[4] Nothing can be found specifically about 6.1231–6.1232 in Landini (2007), Potter (2009) or Kuusela and McGinn (2011). The doctrine of illustrating necessity (namely the doctrine, expressed at 6.12, that tautologies show the "formal" properties of language and the world) upstages the issue of the actual meaning of logical validity. However, the doctrine of showing—together with the Hauptsatz (5.4) that there are no such things as "logical constants"—is but one side of a two-pronged reflection. The other, more positive side lies precisely at 6.1231. An indication of that, as noticed by Anscombe,[5] is the fact that Wittgenstein mentions certainty instead of necessity at 5.525. Logical necessity is not to be dispensed with. However, Anscombe does not say much more. The purpose of this chapter is to analyze Wittgenstein's characterization of logical generality and to show how the Tractarian concept of logical validity distances itself from Russell's in a positive way.

There are at least three reasons why the Tractarian concept of essential general validity has been relatively sparsely discussed. Firstly, as just mentioned, the doctrine of showing, and the mainly negative philosophy of logic advocated in the *Tractatus*, have been the main focus of attention.[6] Secondly, Russell's theory of types has seemed to be the main target of Wittgenstein's criticism in the text being considered. Finally, and most importantly, many commentators have considered that 6.1231–6.1232 is really about logical *necessity*—about the kind of necessity peculiar to logical truths as tautologies:

> Logical propositions need not be general, since a tautology containing names—an 'ungeneralized proposition' (6.1231*c*)—may be valid in virtue of its form. And when logical propositions are general, they are not 'accidentally' valid, do not *happen* to be true of everything, but have 'essential validity', because they treat the 'formal' aspects of the world (6.12*a*).[7]

All three above-mentioned reasons are often associated: Wittgenstein would attack the theory of types in *Principia Mathematica* because the latter epitomizes the logical perspective that tries to express the formal features of the world as general facts through propositions which, although generalized, cannot be but contingently true, if true at all. Putting forward essential general validity and setting it apart from accidental generality would amount to criticizing the endeavor to describe as a fact, however general, what can only be shown as part of the scaffolding of the world.

> As known, the opposition between essential and accidental general validity of logical propositions is elaborated in the *Tractatus* in overt polemic with Russell's conception of logic. In effect, the above-mentioned inadequacy of the logicist reduction of arithmetic follows as a simple corollary from what Wittgenstein considered one of the main flaws of Russell's theory of types: among the axioms of the theory, there are propositions that,

[4]See Hacker (1989, pp. 38 and 47).

[5]Anscombe (1959, p. 158).

[6]As an example, see Stenius (1960, p. 201).

[7]Black (1964, p. 326).

> despite their complete generality, do not satisfy the basic requirement of logical validity, i.e. truth in all possible worlds. […] According to Wittgenstein, an essential general validity is one that does not depend whatsoever on the real configuration of the world, on the particular arrangement of objects which actually occurs.[8]

This reading is followed by Roger White.[9] It is mistaken, however. First, 6.12 does not say that tautologies depict or even show the formal aspects of the world, but that certain propositions being tautologies shows the formal properties of the world. So the tautological nature of certain propositions cannot be explained in terms of formal properties, let alone necessary features of the world: It is the other way around.[10] Moreover, and above all, even in its opposition to accidentality and fortune (at 6.1232), the concept of the general validity of logical propositions aims to pinpoint first and foremost, before a certain kind of necessity, a certain kind of *generality*. When Wittgenstein says (at 6.1231) that being tautological does not imply being generalized, he only means that a tautology does not have to be universally quantified, and actually that the generality of a tautology has nothing to do with quantification. This is borne out by the examples given by 6.1232, and rightly acknowledged by Marie McGinn:

> One of the main themes of Wittgenstein's reflections on the propositions of logic in the *Notebooks* is the attempt to make clear the distinction between the propositions of logic and fully generalized, material propositions in which all the constants have been replaced by variables. Clarification of this distinction is fundamental to Wittgenstein's overall aim to make clear that the sort of generality that belongs to the propositions of logic is quite distinct from the merely accidental generality of general empirical propositions. The generality that characterizes logic has nothing to do with general truth, but with the generality of logical form, that is, with something that abstracts from all content.[11]

---

[8]Frascolla (1994, pp. 38–39).

[9]"What is peculiar about a truth of logic is that you can tell that it is true from the symbol alone, so that the proposition is true regardless of the way the world is, i.e., *necessarily* true." (White 2006, p. 106).

[10]Cora Diamond ("Throwing Away the Ladder: How to Read the *Tractatus*", Chap. 6 of Diamond 1991) rightly insists against Hacker that logical necessity has nothing to do with "features of reality" whatsoever, but pertains to linguistic *constructions*. She writes, about 6.37 and 6.375: "Logical necessity is that of tautologies. It is not that they are true because their truth conditions are met in all possible worlds, but because they have none. 'True in all possible worlds' does not describe one special case of truth conditions being met but specifies the logical character of certain sentence-like constructions formulable from sentences." (Diamond 1991, p. 198) However, she adds: "But the remark that there is only logical necessity is itself ironically self-destructive. […] In so far as we grasp what Wittgenstein aims at, we see that the sentence-form he uses comes apart from his philosophical aim. If he succeeds, we shall not imagine necessities as states of affairs at all. We throw away the sentences about necessity; they really are, at the end, entirely empty." (ibid.) If one were to follow Diamond's reading, one should conclude, not only that logical necessity is not descriptive, but that it is indescribable. Yet it entirely remains to understand what the necessity of tautologies consists in exactly.

[11]McGinn (2006, p. 59).

Admittedly, Wittgenstein's target is partly Russell's proposal simply to replace necessity with universal validity, notably in the paper "Necessity and Possibility," drawn from a talk that Russell gave at the Oxford Philosophical Society in October 1905:

> It is possible to regard a proposition as *necessary* when it is an *instance* of a type of propositions all of which are true. For example, "Socrates is either a man or not a man" may be called necessary on the ground that the statement remains true if we substitute anything else in place of Socrates. […] To make our definitions of necessity and possibility precise, in this theory, it is natural to regard necessity and possibility as not attaching to propositions, but to propositional functions, that is, to propositions with an indeterminate subject. Thus we may define as follows:
>
> "The propositional function 'x has the property ϕ' is *necessary* if it holds of everything; it is *necessary throughout the class u* if it holds of every member of u."[12]

Thus, Russellian general validity is always the general validity of some propositional function $\varphi\hat{x}$, which itself amounts to the truth of the corresponding universally quantified proposition $(x).\ \varphi x$.[13]

However, Russell-averse as it may be, 6.1231–6.1232 does *not* seek to vindicate necessity against Russell's rendering of it. Indeed, the essential validity put forward in the *Tractatus* remains an essential *general* validity. Wittgenstein's purpose at 6.1231–6.1232 is *not* to rebuke Russellian logical generality in favor of logical necessity understood as tautologyhood. On the contrary, it is to bring out the right notion of logical generality against the quantificational generality on which Russell relies exclusively.

There is a traditional assumption that generality and necessity can both be equally considered as criteria of logical validity, and are, as such, equivalent. In particular, Kant writes in his *Logic*:

> If […] we put aside all cognition that we have to borrow from *objects* and merely reflect on the use just of the understanding, we discover those of its rules which are necessary without qualification, for every purpose and without regard to any particular objects of thought, because without them we would not think at all. […] And from this it follows at the same time that the universal and necessary rules of thought in general can concern merely its *form* and not in any way its *matter*. Accordingly, the science that contains these universal and necessary rules is merely a science of the form of our cognition through the understanding, or of thought. And thus we can form for ourselves an idea of the possibility of such a science, just as we can of a *universal grammar*, which contains nothing more than the mere form of language in general, without words, which belong to the matter of language.

---

[12]Russell (1992a, pp. 518–519).

[13]Gregory Landini (see Landini 2007, Appendix B) holds that Russell defines as necessary any true fully generalized proposition in the sense of its first-order *and second-order* universal generalization. For instance, $\varphi x \supset_x \psi x\,.\,\varphi a : \supset .\ \psi a\ \ (1_R)$ is Russell-necessarily true because its full universal generalization $(\varphi)(\psi)(y) : .\ \varphi x \supset_x \psi x\,.\,\varphi y : \supset .\ \psi y\ \ (2_R)$ is true. One possible objection is that necessity then becomes the feature of a sentence, i.e., of the particular linguistic expression of a proposition. For suppose Px is $\varphi x \supset \psi x$. Then (1_R) is equivalent with $(x)Px\,.\,\varphi a : \supset .\ \psi\ \ (1'_R)$, whereas (2_R), which is true, is not equivalent with the full universal generalization of $(1'_R), (P)(\varphi)(\psi)(y) : .\ (x)Px\,.\,\varphi y : \supset .\ \psi y\ \ (2'_R)$, which is false.

> Now this science of the necessary laws of the understanding and of reason in general, or what is one and the same, of the mere form of thought as such, we call *logic*.[14]

As one can see, Kant indiscriminately describes logic by its formal generality and by its absolute necessity. Kant and Russell obviously do not think of the generality of logic in the same way. Kant sees logic as a canon (a body of rules), whereas Russell sees it as a science (a system of truths). As a consequence, the generality of logic, to Kant, lies in the applicability of logical rules independently of the content of the judgments at stake. In contrast, to Russell, it lies in the unrestrictedness of logical truths as pertaining to all entities whatsoever. To Kant, logic is contentless, whereas to Russell it has a content with the widest scope. As is well known, Kant distinguishes between transcendental logic and formal logic, and the validity of the transcendental principles of pure understanding cannot be established for itself, "but always only indirectly through relation of these concepts to something altogether contingent, namely, possible experience."[15] However, transcendental logic is no less general than formal logic, because the domain ruled by formal logic cannot be seen as extending the domain ruled by transcendental logic. No domain properly speaking can embrace and enlarge the totality of possible experience. In that setting, the generality of formal logic has to be reckoned through something other than its universality, namely: Formal logic states conditions of possibility of thought as such. To Kant, necessity is a way to account for formal generality from within a non-formal conception of logic. Russell's strategy goes the other way round: The generality of logic allows Russell to express the special validity of logical truths from within an anti-psychologistic conception of logic, i.e., despite "the fact that the division of judgments into necessary, assertorical, and problematic is, in the main, based upon error and confusion."[16] Unrestricted generality is Russell's way of reckoning the existence of logical truths without committing himself to some kind of necessity that could not be captured formally.[17]

In 6.1231–6.1232, Wittgenstein certainly follows the philosophical tradition, starting with Kant, which associates the formal generality of logic with its particular validity. The *Tractatus* does *not* aim to depart from this tradition. It rather deepens it so as to connect it back to its genuine root and to clarify the combination of generality and necessity peculiar to logical truths. Explaining this combination is the main concern of the rest of this chapter.

---

[14]Kant (1992), "The Jäsche logic," p. 528 (I, Ak. ix, 12).

[15]Kant (1929, p. 592; A737/B765).

[16]Russell (1992a, p. 508).

[17]About Russell's wariness about modal notions, which does not amount to "wholesale eliminativism," see Shieh (2011, pp. 3–4).

Essential Generality

"Essential" general validity has first to do with generality. On that score, the canonical text in the *Tractatus* is 5.501, which distinguishes three ways of determining the values of a propositional variable:

> We *can* distinguish three kinds of description: 1. direct enumeration, in which case we can simply substitute for the variable the constants that are its values; 2. giving a function *fx* whose values for all values of *x* are the propositions to be described; 3. giving a formal law that governs the construction of the propositions, in which case the bracketed expression has as its members all the terms of a series of forms.[18]

A series of forms (4.1252) is a series of terms (or propositions) that is ordered by "internal relations" shown in the symbols for the terms, so that each term (except the first) can be deduced from the previous one according to formal features expressing the same formal law. The example given at 4.1273 is $aRb, (\exists x) : aRx \,.\, xRb, (\exists x, y) : aRx \,.\, xRy \,.\, yRb, \ldots$; The general term of a series of forms is written: $[a, x, O'x]$ (5.2522).

Given a propositional function ξ, Wittgenstein writes '$\bar{\xi}$' for the class of all its values. This shift is in no way automatic. What Wittgenstein calls a variable is precisely the determination of a class of propositions as a source of generality. The generality lies entirely in the mode of determination of such a class that has to be described and thus unfolded before it can be ranged.

Wittgenstein apparently considers three main ways to present a domain of generality. The trichotomy is superficial, however. Indeed, direct enumeration is not a mode of determination among others, but the ideal original datum in comparison to which any mode of determination can appear as a particular mode, a particular way of producing a class of objects. It becomes meaningless, however, in the infinite case, which is the main stake of logical generality. The use of a propositional function is the second mode of determination that Wittgenstein reckons with. However, the only determination that such a device (the sign of a function) is able to carry out is purely nominal—unless a magical "range of significance" is supposed to be attached to any function, and given along with it: This is actually Russell's last resort in "Mathematical logic as based on the theory of types."[19] The functional notation of the form *fx* used by Frege and Russell is also criticized at 5.521:

[18] Wittgenstein (1974, p. 49).

[19] "[…] every propositional function has a certain *range of significance*, within which lie the arguments for which the function has values." (Russell 1956, pp. 72–73)

> 5.521. I dissociate the concept *all* from truth-functions.
>
> Frege and Russell introduced generality in association with logical product or logical sum. This made it difficult to understand the propositions '$(\exists x).fx$' and '$(x).fx$', in which both ideas are embedded.[20]

The use of functional symbols completely glosses over the task of generating and specifying the class of all its admissible arguments. This task is generally discharged in mathematics (in mathematical analysis) by the theoretic framework in which the "domain of definition" of each function is amenable to complete determination. Mathematical analysis is but a particular case, however, and nothing guarantees such a determination in the general case, i.e., in the context of logic.[21] The sole genuine mode of determination of the values of a variable consists in a series of forms. In that case, the class of all values is properly generated, following the operation "that produces the next term out of the proposition that precedes it" (4.1273). The general term of a series of forms is a variable, and any genuine variable is given through a series of forms. Quantificational generality $((x).fx)$ relies on a function notation (fx, or $f(x)$) that does not explain by itself, within its own symbol, how the values of x can be determined, and thus how it makes sense.

The function $f\hat{z}$ can actually be rephrased as the following degenerate series of forms:[22]

$$[(fa,\ z),\ (fx,\ y),\ (f[y/x],\ z)].$$

This rephrasing allows one to see what is right and what is wrong in unspecified quantification. Indeed, the difference between

$$aRb, \exists x(aRx\,.\,xRb), \exists(x,y)(aRx\,.\,xRy\,.\,yRb), \ldots \tag{1}$$

and

$$\varphi(a), (\varphi(a), \varphi(x)), (\varphi(a), \varphi(x), \varphi(y)), \ldots \tag{2}$$

is obvious. In (1), one has to design ever new variables, so there is no reason why this should be refused in the case of (2), and this is enough to bestow a minimal meaning on the rephrasing above of $f\hat{z}$ as a degenerate series of forms. However, 'x' and 'y' have the status of variable in (1), which becomes completely unclear in (2), so that it is impossible to know how to continue. In other words, the variable involved in a propositional function written fx or $f\hat{z}$ (and taken as such) is only the sign of a variable, not a true variable symbol.

In Wittgenstein's view, the main and probably only logically correct expression of generality resides in the iterability of the operation that underpins a series of

[20] Wittgenstein (1974, p. 51).

[21] It should be noted that the phrase "propositional function" occurs only once in the whole *Tractatus*, at 5.521.

[22] Cf. Wittgenstein (1978, p. 453).

forms. The logically correct variable symbols are the variables whose values are specified by a formal law—"form-series variables," as Thomas Ricketts calls them.[23] The logically correct generality is the generality of the general term of a series of forms, which we shall call *form-series generality*. This is confirmed by Wittgenstein's *Notebooks*[24] and by his later remarks on the difference between totalities and systems.[25] Form-series generality is essential precisely to the extent that it is governed "by an internal relation" (4.1252) and resides "in the symbol itself" of the series (5.525, the connection with 6.113 being obvious). The exact opposite is the "*accidental* generality" carried by the theory of classes (6.031), and to which corresponds the domain of values that quantification merely presupposes. Quantificational generality leaves the propositional variable ξ (the class of all its values) up in the air. In contrast, form-series generality pertains by definition to a form-series variable whose class of values is produced through an operational rule built into the symbol of a series of forms.

Of course, quantificational generality is not dismissed at all by the *Tractatus*.[26] As a notational system, quantification is irreproachable: This is what is to be gathered from 4.0411. The point is elsewhere, in the way in which a notation can be filled in with an actual meaning. In this respect, quantification is certainly a kind of mixed form straddling ordinary language and mathematical language. Quantification, applied to propositional functions, is a very versatile tool which allows one to formalize both "There are two objects which are laid on my desk" and "There are two objects which are a solution of the equation $x^2 = 1$."[27] As such, the quantificational notation is a powerful tool to uniformize the various aspects of generality in language. As a counterpart of this virtue, quantificational variables cannot by themselves be more than a notational device: they do not show their range by themselves. This is not a problem in the context of ordinary language, or if the framework of the theory of classes is endorsed. However, this cannot do as an exact logical representation of how *logical* generality actually works; This cannot do as an exact analysis of the generality of logical propositions. In the case of an empirical propositional function, it is part of the meaning of the latter that its extension is determined factually. However, in the case of a logical or a mathematical concept, it cannot be left to the world to determine its extension, i.e., to produce the range of values of the corresponding variable.[28]

The objection of "*accidental* generality" that Wittgenstein raises at 6.031, against the logicist reconstruction of numbers as equinumerosity classes of extensions, can

[23]Ricketts (2013, p. 136).

[24]Wittgenstein (1961, 13.10.14, p. 11).

[25]Wittgenstein (1967, pp. 216–217).

[26]See, in particular, 4.0411, 5.1311, and 5.52.

[27]See Wittgenstein (2001, pp. 125–127 and 171–175). On this point, see Sackur (2005, pp. 133–141).

[28]Wittgenstein (1961, 13.10.14, p. 11): "But let us remember that it is the *variables* and *not* the sign of generality that are characteristic of logic."

be better viewed in that light. Wittgenstein's criticism aims at two targets here: the use of an abstract extensional generality on the one hand, and on the other the resort to propositional functions having as by chance the right logical properties. Let us indeed consider the definition of 0. In Frege's and Russell's view, non-self-identity evidently secures in a logical way the existence of an empty class, and thus the possibility of defining 0. In other words, it is self-evident to both Frege and Russell that the following sentence (s), $(x).x \neq x$, constitutes a logical truth. According to Wittgenstein, it is not, as no concept is able to determine in a logical way the cardinality of its extension. One can conceive of a possible world where each propositional function would be satisfied by at least one object. So, should the variable sign 'x' be a logical term, (s) would be a logical sentence (because all its constituents would correspond to logical notions) and a true sentence, and yet would not be a logically true sentence. This has less to do with the rejection of the identity sign (5.53–5.534) than with a general discussion of logicality, as proved by 5.5352:

> 5.5352. In the same way people have wanted to express, 'There are no *things*', by writing '$\sim(\exists x).x = x$'. But even if this were a proposition, would it not be equally true if in fact 'there were things' but they were not identical with themselves?[29]

Thus, Wittgenstein implicitly confronts logicists with the perspective that the truth of a logical sentence does not imply the logical truth of that sentence. The conclusion to draw is that (s) is not really a logical sentence, because a variable sign such as 'x' is not a logical term as such.[30] Everything hinges on the way a variable is introduced and used: on the existence of rules for its referentiality, i.e., on the corresponding "mode of signification" (in the sense of 3.322). Only a form of series can specify in a principled way the range of values of a variable. Only form-series variables uncover what enables a proposition to be general; This is exactly why they can be described as an "essential" mode of signification, i.e., as a mode of signification based on the "essential features" of generality, in the sense of 3.34–3.341: "Essential features [of the propositional sign] are those without which the proposition could not express its sense."[31] *Essential generality* can be defined as the generality of a proposition expressed by essential features of the symbol of generality. In contrast, quantificational generality is a blank check that expresses its purported sense only superficially and thus qualifies as accidental generality. The main conclusion that can be drawn from 5.501 is that essential generality is fully captured by form-series generality only.

The extensional determination of the range of values of a variable does not make sense as such (except maybe in the case of a finite extension, i.e., in the case of an enumeration) because the reference to an extension, without any rule to range and survey it, does not educe any properly logical specification at all. A variable sign lacks any mode of signification and remains logically undetermined if the task of

[29]Wittgenstein (1974, p. 53).

[30]This is actually, in a way, Tarski's conclusion: The interpretation of a variable changes as one shifts from one "universe of discourse" to another.

[31]Wittgenstein (1974, p. 17).

determining the identity and number of the objects which satisfy some propositional function is delegated to the world instead of residing in the symbol itself. Of course, if one says "There is a pen somewhere on this table," some generality is actually expressed in a perfectly meaningful way. However, such a kind of generality has nothing to do with logical generality—the kind of generality that is expected and required in the case of logical truths.[32] This diagnosis, which goes beyond Wittgenstein's hostility toward set theory, explains the stress put by Wittgenstein on induction throughout his later reflections on mathematics. Inductive generality, drawing on form-series generality, became the paradigmatic figure of logically controlled, non-enumerative generality at the time *Philosophical Grammar* was written[33]: There are, on principle, as many kinds of generality as there are discursive "systems,"[34] but this does not detract from the fact that the iterability of an operation constitutes the core of mathematical generality.[35] This is what can already be found in the *Tractatus*. As underlined at 5.501, the variety of notational systems to express generality remains an important feature of language, which no essential form of generality should completely obliterate. However, there does exist an essential form of generality, namely form-series generality.

So the conclusion is that the "essential" general validity of logical propositions has to have to do with some series of forms. It remains to be seen which.

Essential General Validity

Let us consider $p.p \supset q : \supset . q$. There is no question that this proposition is both generally valid (owing to some sort of schematical generality) and necessarily true as well. It remains true however 'p' and 'q' are replaced with other propositions, and it remains true in any possible situation as well. The real question is how are we to represent these two features, and understand their combination?

6.1231–6.1232 clearly mentions both generality and essential validity (as opposed to "accidental" validity, or truth "as a result of a fortunate accident"). Now, one thing is clear: The essentiality that the *Tractatus* ascribes to the general validity of logic comes not only from an essential (i.e., non-accidental) validity, but also from an essential generality. Essential general validity, indeed, involves some generality, yet it follows from the above that a mere generalized (i.e., universally quantified) proposition cannot have the essentiality required by logical truth:

[32]Wittgenstein (1961, 23.10.14, p. 17): "If the completely generalized proposition is not completely dematerialized, then a proposition does not get dematerialized at all through generalization, as I used to think."

[33]On this topic, see Marion (1998, p. 98 sq.).

[34]Wittgenstein (1978, pp. 458–459): "How a proposition is verified is what it says. Compare generality in arithmetic with the generality of non-arithmetical propositions. It is differently verified and so is of a different kind."

[35]Wittgenstein (1978, p. 457): "What is general is the repetition of an operation."

Generalization cannot mean but a general fact. So the non-accidentality of the general *validity* of *logic* also requires the non-accidentality of the *general* validity of logic. *How, then, to combine essential generality and essential validity so as to get the essential general validity that 6.1232 assigns to logical propositions?*

The question is legitimate, because generality and validity point to different directions. Indeed, validity has to do with truth-functionality, and 5.521 insists precisely on the separation to be made between generality and truth-functionality. This separation is what the operator N is meant to bring out.

Let us consider this latter point. Fogelin[36] has deemed the operator N to be inadequate to indicate the respective scopes of variables and, as a consequence, to capture quantifier alternation (i.e., to express mixed quantified formulae such as $(x)\ .\ (\exists y)\ .\ fxy$ or $(\exists x).\,(y).fxy$). This assessment sparked a very substantial literature, devoted to evaluating the possibility of recovering first-order logic from the operator N.[37] Along with the issue of the operator N's expressive sufficiency, the question has also been raised as to whether the *Tractatus* is committed (by 6.113) to a decision procedure for checking that a given proposition is a logical truth.[38] Although those questions are interesting by themselves, the introduction of N does not purport to replace the syntax of quantification, let alone to improve on it. This is most definitely borne out by 4.0411. The whole rationale of the operator N is to symbolize verifunctionality in the most general way and to distinguish as clearly as possible what 5.521 says that Frege and Russell failed to distinguish clearly, namely generality (i.e., the introduction of a class of propositional values) and verifunctionality (i.e., a certain combination of the corresponding truth values). The logicist functional notation precisely conflates both components: A propositional function is supposed to simultaneously select a domain of arguments, assign a value to each of them, and turn out the logical product (or the logical sum) of all values.[39] In contrast, the notation '$\mathrm{N}(\bar{\xi})$' introduced at 5.501–5.502 is geared to represent, in the most general and perspicuous way, the combination of two distinct factors represented as distinct: the determination of a variable, which consists in the selection of a system of propositions, as represented by $\bar{\xi}$; and all the possibilities of truth-functional calculus, as represented by N. The separation of both factors is the condition of their *combination*, as opposed to their confusion.

Form-series generality and truth-functional validity thus refer to independent propositional structures. Essential general validity, however, corresponds precisely to the case where they become indistinguishable. How to understand this?

Indeed, if one stresses the sole generality of a tautology such as $p\ .\ p \supset q : \supset\ .\ q$, one is led to describe it in terms of quantificational generality (understood

[36]Fogelin (1976, particularly pp. 78–79).

[37]See Geach (1981, pp. 169–170), Soames (1983, p. 578), McGray (2006), Landini (2007 pp. 136–138), and Rogers and Wehmeier (2012).

[38]On this question specifically, see McGray (2006, pp. 161–168).

[39]This point is well brought out in Sackur (2005, pp. 129–137). See also McGinn (2006, pp. 231 and 235–236).

substitutionally or otherwise), in the sense of its being made fully explicit as $(\phi)\ .\ (\psi) : \phi\,.\,\phi \supset \psi\ :\supset\ .\ \psi$. However, this is clearly incompatible with what Wittgenstein has in mind:

> […] Wittgenstein believes that we must be careful to distinguish the general proposition of logic from generalizations of material propositions. On his view, construing $(p)(p \vee -p)$ as a substantive general truth about logical objects, obscures this distinction.[40]

On the other hand, stressing the sole truth-functional validity of a tautology amounts to considering the latter as a truth that remains true in every possible world, and this is equally misleading, because it clearly presupposes (and only rephrases in a figurative way) what is to be analyzed. A logical proposition is a proposition whose truth can be established on the basis of its symbol alone. This has nothing to do with evaluating it in various possible worlds so as to bring out some kind of invariance. By way of analogy, logical truth is to truth across all possible worlds what "direct reference" is to "obstinately rigid designation," to use David Kaplan's terminology. An obstinately rigid designator is a term which designates the same individual object in every possible world (whether that object exists in that world or not). Kaplan takes a directly referential expression to be rigid in the deeper sense that its referent, once determined, is taken to be fixed for all possible circumstances because it is the propositional component itself:

> For me, the intuitive idea is not that of an expression which *turns out* to designate the same object in all possible circumstances, but an expression whose semantical *rules* provide *directly* that the referent in all possible circumstances is fixed to be the actual referent.[41]

Kaplan explains the rigidity of proper names as a *consequence* of their being directly referential:

> If the individual is loaded into the proposition (to serve as the propositional component) before the proposition begins its round-the-worlds journey, it is hardly surprising that the proposition manages to find that same individual at all of its stops, even those in which the individual had no prior, native presence. The proposition conducted no search for a native who meets propositional specifications; it simply 'discovered' what it had carried in.[42]

In Kaplan's view, direct reference is the deep semantic structure explaining (obstinately) rigid designation. In the same way, tautologyhood represents in the

[40]McGinn (2006, p. 60. See also p. 248): "A particular instance of a proposition of this form [namely, $(p \vee -p)$] is not a logical truth in virtue of being a substitution instance of a general logical law, $(p)(p \vee -p)$, but simply in virtue of the way it is constructed, that is, simply in virtue of its having the form $(p \vee -p)$." Ricketts (2013, p. 127) also insists on the distinction to be made between a relation between entities on the one hand, and the construction of a logical structure such as a material conditional on the other. Consequently, logical generality cannot consist in the substitutability of entities, but only in the generality of a construction, i.e., in the iterative generativity of a truth-operation.

[41]Kaplan (1989, p. 493).

[42]Kaplan (1989, p. 569).

Tractatus the deep symbolic structure that explains the invariance in truth value of a logical proposition.[43] Just as it is guaranteed in advance of evaluation that a directly referential term has one and the same referent in all possible worlds, analogously it is guaranteed in advance that a logical proposition remains true across all possible worlds. Logical propositions pertain to everything that is essential to a proposition's being true or false,[44] so the notion of possible world (defined as a certain assignment of truth values) is derivative of that of logical truth—which is why one cannot resort to the former to explain the latter.

Thus, although generality and truth-functional validity correspond to independent structures, the special validity peculiar to logical propositions requires more than their external combination, as in the case of regular general propositions: It requires their coincidence. So there remains only one option: The series of forms underpinning the form-series generality of a given logical truth has to be identical with the series of forms underpinning truth-functionality in general. Let us try to expand upon this idea.

At 6, Wittgenstein identifies the general form of a proposition with the general form of a truth function, as given by the series of forms $[\bar{p},\ \bar{\xi}, \mathrm{N}(\bar{\xi})]$. As emphasized by 6.001, this general form means that any proposition can be exactly identified with the series of forms (starting from certain elementary propositions) that leads to it. In the same way, a tautology *is* nothing else but the drawing of the "truth-diagram"[45] showing that no combination of truth values can lead to the pole written F, as shown at 6.1203 in the case of $\sim(p.\sim p)$. The general validity of a tautology follows from its method of constructing only. In the case of $p\ .p \supset q : \supset\ .q$, p and q stand for any propositions: They act as mere indices of truth values. However, this holds exclusively in the context of the tautologyhood of the whole proposition. The general validity of the tautology certainly cannot be rendered by $(p)\ .\ (q) : .p\ .\ p \supset q : \supset\ .q$, because the generalization over p and q makes sense only after the tautologyhood of the proposition has been granted, precisely. As soon as the tautologyhood of the proposition is established, however, it becomes obvious that p and q can be replaced by whatever propositions one may wish. The general validity of a tautology is thus internal to the propositional symbol that one wishes to generalize over. That is the reason why it eludes any generalization (in the form of a universal quantification), as claimed by 6.1231. In other words, the truth-diagram of $p\ .p \supset q : \supset\ .q$ is integral to the full symbol of that tautology, and its method of constructing shows, as an essential feature of it, how 'p' and 'q' are thereby neutralized. This is the "zero-method" mentioned by 6.121. Hence, the general validity of a tautology is essential both because its validity is

[43]This point dovetails with Diamond's remarks quoted in Footnote 10. As Sanford Shieh summarizes: The reason why a tautology is true in a special way "is not because it describes a special type of invariably obtaining situations. […] it is the nature of linguistic representation, rather than features of the world or of all possible worlds, that makes tautologies true." (Shieh 2011, p. 3)

[44]As Wittgenstein puts it in his *Notebooks*: "The logic of the world is prior to all truth and falsehood." (Wittgenstein 1961, 18.10.14, p. 14)

[45]Potter (2009, pp. 160–164).

internal to its symbol and because its generality is internal to its validity (being neither its ground nor its external outcome).

To sum up, in the symbol-directed approach to both logical generality and logical necessity that is pursued by the *Tractatus*, the generality as well as truth-functional invariance of a logical proposition are built into its symbol. As a result, it is meaningless to express the generality through some universal quantification or some substitution rule. The universal quantification or the substitution rule can be but derivative and inaccurate ways of speaking. For the same reason, it is meaningless to describe the proposition as remaining true however the truth-value assignment is changed, as if the proposition were attached to a particular original assignment to begin with.

The series of forms $[\bar{p},\ \bar{\xi},\ \mathrm{N}(\bar{\xi})]$ is not exhausted by its role of presenting the general form of a proposition as a truth function. Landini objects to Anscombe's identification of the general form of a truth function, $[\bar{p},\ \bar{\xi},\ \mathrm{N}(\bar{\xi})]$, with the general term of a consecutive series of truth functions, $[\bar{p},\ \mathrm{N}^{n}(\bar{p}),\ \mathrm{N}^{n+1}(\bar{p})]$:[46]

> Wittgenstein's assertion that *all* truth-functions can be reached by "successive applications" of the N-operator does not require Anscombe's identification of his notion of "the general form of a truth-function" with the notion of the *general term* of a *consecutive* series of truth-functions. When Wittgenstein said that every truth-function is the result of successive applications of the N-operator, he may have simply meant that the N-operator is expressively adequate.[47]

Accordingly, the general term corresponding to the general form of a proposition is not meant to represent all propositions as belonging to a single ordered series (with *all* elementary propositions as its basis) but rather, more flexibly, any particular truth function (with only the relevant subset of all elementary propositions as its basis: the set of those elementary propositions that occur in its constructional history).[48] This fits Wittgenstein's use of the bar (at 5.501) as standing for the determination of a certain selection: a certain selection of the bases in the case of $\bar{p}$, or a certain selection of propositional values in the case of $\bar{\xi}$.[49]

This flexible adjustment of the notation $[\bar{p},\ \bar{\xi},\ \mathrm{N}(\bar{\xi})]$, depending each time on the particular truth-functional setting at stake, allows one to understand how the general form of a truth function can coincide with the series of forms ruling the generality of a particular logical proposition. Indeed, given a logical proposition $\Theta(\bar{p})$ (for instance, $\bar{p} = \{P, Q\}$ in the case of $\Theta = P\,.\,P \supset Q\ :\supset\ .\,Q$), one can express both its essential generality and its essential validity by writing:

---

[46]Anscombe (1959, p. 132).

[47]Landini (2007, p. 143). See also Ricketts (2013, p. 140): "[…] the general form of sentences has an intrinsically schematic character. […] The general sentence-form is rather a scheme for the construction of any sentence: it is the most general form for the construction of truth-functions of elementary sentences."

[48]In the series $[\bar{p},\ \bar{\xi},\ \mathrm{N}(\bar{\xi})]$, contrary to what McGinn claims (McGinn 2006, p. 234), $\bar{p}$ does *not* have to be the totality of all elementary propositions.

[49]This point is recalled by Landini (2007, p. 142).

$$\Theta([\bar{p}, \bar{\xi}, \mathrm{N}(\bar{\xi})]) = [\Theta(\bar{p}), \Theta(\bar{\xi}), \Theta(\mathrm{N}(\bar{\xi}))].$$

The left-hand side of this identity stresses the truth-functional validity of Θ, whereas its right-hand side stresses the form-series generality of Θ. Indeed, the expression on the left provides the truth-functional calculation that takes place when elementary propositions selected within the basis $\bar{p}$ of Θ are replaced with their respective negations, in other words when the assignment of truth values to elementary propositions of Θ is changed, and generally when the truth value of Θ is considered for all possible truth-value assignments. As to the expression on the right, it implements the general form of a proposition within Θ, which corresponds to the fact that the basis of Θ can be replaced by any appropriate selection of propositions: all occurrences of '*P*' replaced by any proposition ϕ and similarly for all the other members of $\bar{p}$. This is actually what the end of 6.124 conveys: "If we know the logical syntax of any sign-language, then we have already been given all the propositions of logic,"[50] which means conversely that any proposition of logic already carries the whole logical syntax of all propositions.

Of course, both sides of the identity above correspond to the same thing, namely the general validity that 6.1232 seeks to bring out. The series of forms that gives the general form of a truth function and corresponds to the truth-functional invariance peculiar to tautologies is at the same time the series of forms that grounds the essential generality peculiar to logical propositions. This coincidence of the two series of forms is finally what explains why the essential generality and the essential validity of the propositions of logic merge into a single essential general validity.

It is now possible to get back to the starting point of sections 5.501 and 5.521, which convey a criticism of both Frege and Russell. Two distinct kinds of generality run in parallel in Frege's *Begriffsschrift*: quantificational generality, as expressed by means of individual variables, and schematic generality, as expressed by means of schematic letters for judgeable contents (any particular judgeable content being substitutable for a schematic letter). Both kinds, however, are combined, as illustrated for instance by the proof of judgment (93), where both substitutions $(a \mapsto \mathfrak{f}, f(\Gamma) \mapsto \Gamma(y))$ are carried out.[51] Wittgenstein implicitly objects to this ambiguity. There are two symmetric ways to clear it up: either by considering schematic letters as implicit variables (this is the Russellian way, with the delicate distinction between *entity-variation* and *meaning-variation*)[52] or by tracing propositional variables back to schematic form-series generality: This is the Tractarian way.

The second criticism that can be found at 5.521 is about Russell. Wittgenstein's target is not only the quantificational symbolism of *Principia Mathematica* but also,

[50] Wittgenstein (1974, p. 63).

[51] Frege (1972, p. 183).

[52] See Russell (1992b, pp. 360 ff).

maybe primarily, the substitutional theories that Russell developed between 1905 and 1908 and whose main published expression is "On 'Insolubilia' and Their Solution by Symbolic Logic." Russell's framework is based on substitution, i.e., on the replacement of some constituent of a given proposition by another entity. For a proposition p and a constituent a of p, the expression '$p/a;b!q$' is taken as primitive, and means that q results from replacing a by b in p. If p is a proposition, $(x) : p/a;x!\,q\,.\&\,.\,q$ then means that a replacement of a by x in p results in a true proposition, for any x whatsoever.[53] The main point of the Tractarian criticism here is that the formal substitutability of b for a in p indicates the substitutability of any entity for a, and thus a variable argument: Generality is already present. In Russell's substitutional logic, the substitution of the entity variable x for a has exactly the same status as the substitution of b for a. However, in Wittgenstein's view, the generality of x substituted for a already lies in the very substitutability of x for a. Russell makes it appear as if the generality carried by x were magically attached to x independently of this substitutional context and could simply be added to the operational apparatus of substitutions. Against this presentation, Wittgenstein claims that the substitutability of expressions (propositions) to others already appeals to some form of generality: ultimately, to form-series generality.

In conclusion, the essentiality of the general validity that the *Tractatus* attributes to logic is finally explained. This validity is described by the *Tractatus* as "essential" because logical generality and logical validity are both thought of as being essential, and as being essential *to each other*. Essential generality is the generality expressed by essential features of a propositional symbol (3.34), and corresponds to what has been called form-series generality (5.501). Essential validity is the validity of a proposition whose truth can be determined by inspection of its symbol alone (6.113), and corresponds to tautologyhood (5.525). Essential generality and essential validity go together because they rely on the *same* series of forms, the one that presents the general form of a truth function. Owing to that form, which is a variable whose values are all the propositions, the elementary propositions involved in the construction of the truth-diagram of a tautology acquire by the same token a schematic generality that makes any proposition substitutable for each of them, and makes it possible in return to recognize a logical form as such.

References

Anscombe, G. E. M. (1959). *An introduction to Wittgenstein's Tractatus*. London: Hutchinson University Library.

Black, M. (1964). *A companion to Wittgenstein's Tractatus*. Cambridge: Cambridge University Press.

Diamond, C. (1991). *The realistic spirit. Wittgenstein, philosophy and the mind*. Cambridge, MA: The MIT Press.

[53] See Russell (1973, pp. 200–201).

Fogelin, R. J. (1976). *Wittgenstein*. London, Henley and Boston: Routledge & Kegan Paul.
Frascolla, P. (1994). *Wittgenstein's philosophy of mathematics*. London: Routledge.
Frege, G. (1972). *Conceptual notation and related articles* (T. W. Bynum, Translated and edited with a Biography and Introduction). Oxford: Clarendon Press.
Geach, P. T. (1981). Wittgenstein's operator N. *Analysis, 41*, 168–170.
Hacker, P. M. S. (1989). *Insight and illusion. Themes in the philosophy of Wittgenstein*. Oxford: Oxford University Press (revised ed.).
Kant, I. (1929). *Critique of pure reason* (N. K. Smith, Trans.). London: Palgrave Macmillan.
Kant, I. (1992). *Lectures on logic*. In J. Michael Young (Ed. Trans.). Cambridge: Cambridge University Press.
Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan (Chap. 18)* (pp. 465–563). Oxford: Oxford University Press.
Kuusela, O., & McGinn, M. (2011). *The Oxford handbook of Wittgenstein*. Oxford: Oxford University Press.
Landini, G. (2007). *Wittgenstein's apprenticeship with Russell*. Cambridge: Cambridge University Press.
Marion, M. (1998). *Wittgenstein, finitism and mathematics*. Oxford: Clarendon Press.
McGinn, M. (2006). *Elucidating the Tractatus*. Oxford: Clarendon Press.
McGray, J. W. (2006). The power and the limits of Wittgenstein's N operator. *History and Philosophy of Logic, 27*(2), 143–169.
Morris, M. (2008). *Wittgenstein and the Tractatus logico-philosophicus*. London and New York: Routledge.
Potter, M. (2009). *Wittgenstein's notes on logic*. Oxford: Oxford University Press.
Ricketts, T. (2013). Logical segmentation and generality in Wittgenstein's *Tractatus*. In P. Sullivan & M. Potter (Eds.), *Wittgenstein's Tractatus. History and interpretation* (Chap. 7, pp. 125–142). Oxford: Oxford University Press.
Rogers, B., & Wehmeier, K. (2012). Tractarian first-order logic: Identity and the *N*-operator. *Review of Symbolic Logic, 5*(4), 538–573.
Russell, B. (1956). Mathematical logic as based on the theory of types. In R. C. Marsh (Ed.), *Logic and Knowledge* (pp. 59–102). London: George Allen and Unwin.
Russell, B. (1973). On 'Insolubilia' and Their Solution by Symbolic Logic. In D. Lackey (Ed.), *Essays in analysis* (pp. 190–214). George Allen and Unwin: London [English Translation of "Les paradoxes de la logique," *Revue de Métaphysique et de Morale* 14/5 (1906), pp. 627–650].
Russell, B. (1992a). Necessity and possibility. *Collected papers* (Vol. 4 (Chap. 22), pp. 507–520). London: Routledge.
Russell, B. (1992b). On Fundamentals. *Collected papers* (Vol. 4 (Chap. 15), pp. 360–413). London: Routledge.
Sackur, J. (2005). *Formes et faits. Analyse et théorie de la connaissance dans l'atomisme logique*. Paris: Vrin.
Shieh, S. (2011). In what way does logic involve necessity? In *Contemporary Tractatus: Wittgenstein's Tractatus and Tractarian themes in philosophy today*. Talk given on March 4, 2011 on the occasion of the 3rd Annual Conference at Auburn University.
Soames, S. (1983). Generality, truth functions and expressive capacity in the *Tractatus*. *Philosophical Review, 92*, 573–589.
Stenius, E. (1960). *Wittgenstein's Tractatus. A critical exposition of its main lines of thought*. Oxford: Blackwell.
White, R. M. (2006). *Wittgenstein's Tractatus logico-philosophicus: A reader's guide*. London and New York: Continuum.
Wittgenstein, L. (1961). Notebooks 1914–16. In G. H. von Wright & G. E. M. Anscombe (Eds.), (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.
Wittgenstein, L. (1967). *Fr. Waismann: Wittgenstein und der Wienerkreis. Aus dem Nachlass herausgegeben von B. F. McGuinness*. Oxford: Basil Blackwell.

Wittgenstein, L. (1974). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuinness, Trans.). Routledge: London and New York.
Wittgenstein, L. (1978). *Philosophical grammar*. In R. Rhees (Ed.), (A. Kenny, Trans.). Berkeley, Los Angeles: University of California Press.
Wittgenstein, L. (2001). *Wittgenstein's Lectures (A. Ambrose, Ed.). Cambridge, 1932–1935*. Prometheus Books, Amherst (New York).

Author Biography

Brice Halimi is an Associate Professor in the Philosophy Department at Paris West University, France. His recent publications include a book on the relationship between universality and necessity since Kant, as well as several papers on Russell. His current research belongs to logic and philosophy of mathematics, and bears in particular on the connections between logic and the traditional branches of mathematics.

Reconstructing a Logic from Tractatus: *Wittgenstein's Variables and Formulae*

David Fisher and Charles McCarty

Introduction

There are large disagreements and even misapprehensions about the logic proposed in Wittgenstein's **Tractatus Logico-Philosophicus** (1984) of 1921. Some interpreters, e.g., the Hintikkas in their (Hintikka and Hintikka 1986), endorse the notion that the Tractarian Wittgenstein embraced a logic wholly or largely at one with conventional, finitary propositional and predicate logics in unison with the now familiar semantics due to Alfred Tarski and students. Others prefer the idea that Wittgenstein offered his readers no particular logic, and certainly no formal system or formal semantics of the sorts common among today's logicians. If one takes the term 'formal system' sensu stricto–so requiring an explicit recursive specification of an abstract formal language on the basis of an explicit finite or primitive recursive alphabet of simple signs–then it is perfectly true that Wittgenstein provides, in **Tractatus**, no formal system. He neither describes nor constructs a fully formal language, even though formal languages and systems over them, as in Frege's **Grundgesetze** (1893/1903), were already in print and familiar to him.

That said, two points should be kept clearly in view. First, an ordinary formal system incorporates at least two subparts: an explicit formal language and, in addition, the specification of a deduction, derivability, or consequence relation. This latter specification is often conveyed syntactically–either via axioms and rules, or by rules alone. Alternatively, the consequence relation can be given semantically in terms of structures and interpretations in such a way that it is not tied down to the details of any particular formal language. Hence, although there may be no

The second-named author is responsible for all translations from Wittgenstein's German.

D. Fisher · C. McCarty (✉)
Indiana University, Bloomington, USA
e-mail: dmccarty@indiana.edu

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_12

instructions in **Tractatus** for generating univocally any particular formal language, there is doubtless sufficient material for a fairly unambiguous specification of derivability or consequence, defined either in terms of truth-tables or in terms of what Wittgenstein calls, at 6.121, the 'zero method' [*Nullmethode*]. That method is a calculus or decision procedure for finitary propositional logic that he explains and exemplifies in remarks 6.12 through 6.122. It is a close cousin, even a homomorphic image, of the later tableaux method (Beth 1955). In the usual case, that of finitary formulae from standard propositional logic, both truth-tables and the zero-method are provably sound and complete relative to the rules of conventional or classical logic. It must be noted that an inviolable restriction to the finitary by no means governs the logic of **Tractatus**. For example, Wittgenstein reminds us, at 6.1203, that his zero-method applies to those cases in which, in the relevant expressions, no generality indication [*Allgemeinheitsbezeichnung*] appears. Also, if the rows of Wittgenstein's truth-tables have his elementary propositions as their guides, those rows may well be infinitely long, since he seems to allow, at 4.21 and 4.211 taken together, there to be infinitely many such propositions.

There is a second crucial point: logics can exist independently of the fine details of any particular formal system. Admittedly, a formal system is an efficient and plain way of indicating a logic, but contemporary formal systems are neither identical with logics nor absolutely required for their communication. Aristotle discovered a logic or logics, as did George Boole. Neither one was the designer of a correlative formal system, strictly understood, although the results of their researches can today be put across in formal terms. Moreover, these two logicians did introduce certain notational innovations for purposes of presentation. For instance, Aristotle may have been the first to employ a letter as a mathematical variable. Formal systems are therefore exceedingly convenient tools in logic, but are not absolutely required for the life of the subject. In the spirit of this recognition, one can allow, even insist, that there is a logic in **Tractatus**, and that it is available to us, but not via any fully realized formal system explicitly contained therein. Wittgenstein's little book does contain descriptions of propositions and their logical forms, sometimes by the use of special notational recommendations, and a relation of consequence over those propositions is indicated.

Now, how ought one to discover and present the (or a) Tractarian logic today? We maintain that, against the (often dark) background afforded by the rest of **Tractatus**, the broad hints contained in three remarks suffice to illuminate the design of a contemporary formal language that is recognizably Tractarian. Here, we carry out this design, conceived all the while as a proposed reconstruction or fragment of a proposed reconstruction–in contemporary terms–of a presentation of the logic of **Tractatus**, and subject to reasonable constraints set upon such reconstructions. One of the constraints is doubtless the requirement that, in the reconstructing, any introduction of technical notions not available in principle to the Wittgenstein of 1921 or earlier be avoided, if at all possible. If not possible, anachronistic technical intrusions must be minimized. In establishing the reconstructed formalism in this article, effort is exerted to employ only those notions and notations that Wittgenstein used at the time and only in the way that he used

them–as far as one can make such determination on the basis of the text. As already asserted, formal languages and systems over them were then to hand, as were recursive definitions of functions and of structures. They were to be found in Frege's writings, in Dedekind's brilliant **Was Sind und was Sollen die Zahlen?** (Dedekind 1888), and elsewhere–together with plainly and self-consciously inductive arguments over those structures. As the reader will see, the sole (mild) technical intrusion into this article will be that of a primitive recursive function of natural numbers. Certainly, that idea was more than implicit in Dedekind's work just cited, but it was not isolated and set out in its full beauty until the publication of (Skolem 1923). (It would seem that Skolem wrote his essay in 1919.)

Finally, we emphasize that this prohibition of technical anachronism in no way governs the metamathematical investigation into the formal system that embodies the proposed reconstruction. Just as one is nowise limited to the astronomy of Kepler's own century in demonstrating that he was mistaken in once thinking there to be only six planets orbiting the sun, so, in uncovering the mathematical properties of the reconstruction, one may call upon techniques and ideas known to contemporary mathematical logic, all the while exercising due care not to project results of such investigations back into the doctrines of **Tractatus** itself. Of course, those mathematical properties may later prove themselves a true aid in the interpretation of that mantic little volume.

Although there is no full and original Wittgensteinian formal language, there are a goodly number of notational suggestions and clear logical form proposals in **Tractatus**. Here are three of the most salient. They are not displayed in the order in which they appear in the book.

> 6 The general form of the truth-function is: $[\overline{p}, \overline{\xi}, N(\overline{\xi})]$.
>
> This is the general form of the proposition.
>
> 4.53 The general proposition-form [*Satzform*] is a variable.
>
> 5.501 (second half of the remark) The values of the variables are prescribed.
>
> The prescription is the description of the propositions that the variable represents.
>
> How the description of the components of the bracket-expression comes about is inessential.
>
> We **can** distinguish three types of description: 1. Direct enumeration. In this case, we can simply put, instead of the variables, their constant values. 2. Specification of a function *fx*, the values of which are, for all values of *x*, the propositions to be described. 3. The specification of a formal law according to which those propositions are constructed. In this case, the components of the bracket-expression are all the components of a form-series.

It is familiar that the *N* operator represents Wittgenstein's generalization of the Sheffer stroke or joint denial. However, that in itself does not seem to provide the central element of the notational proposal. That central element looks to be Wittgenstein's concept of *variable*. The doctrine of variables plays the lead role and, as 4.53 makes plain, affords the sole route to understanding his forms of propositions. To gauge the full import of Remark 6 one must grasp the meaning of

the propositional variable ξ as well as what it means to combine ξ with the overline, so obtaining $\overline{\xi}$, what is called herein its *class*. In **Tractatus**, the variable is always a propositional variable, a variable with propositions as its values exclusively. Wittgenstein provides his most extensive explanation of (propositional) variable in the latter half of 5.501, and, in this writing, we take, at least provisionally, his description of the types of specification as exhaustive, despite the apparent qualification conveyed by the line, 'How the description of the components of the bracket-expression comes about is inessential.' This apparent qualification is read as a principle of extensionality. Informally, it asserts that what truly matters to the propositions Wittgenstein circumscribes is the values of the variables, the members of their classes, and not the precise linguistic terms featuring in the specifications. Hence, at 5.501, Wittgenstein means to convey certain propositional values using direct enumerations, substitution instances of propositional functions, and formal laws. For Wittgenstein to succeed in so conveying, the reader must discern those values. How this comes about is not fully to the point; there may be legitimate alternatives to conveying and to discerning them. This business of alternatives is a matter to which the present exposition will return at several points.

Therefore, the core of the recursive reconstruction is the specification, organized into stages or ranks, of the Wittgensteinian variable. Interlaced with that specification is the ranked hierarchy of W-formulae, representing–modulo logical equivalence–Wittgenstein's propositions [*Sätze*] expressed in the *N*-notation. We define the entire hierarchy over some base language or other; for each appropriate base, there is a proprietary specification of variables and W-formulae. This was, in effect, a feature of Wittgenstein's original design; please remember the principle of extensionality in this connection. In the present writing, the hierarchies over base languages for propositional logic, predicate logic, and arithmetic are objects of attention; a study of such hierarchies arising from finite base languages is left for another time. Lastly, please do not be alarmed that, in the expression 'W-formula' and elsewhere, the term 'formula' from contemporary mathematical logic is employed. Again, Wittgenstein's principle of extensionality guarantees that, to succeed in our task, the detailed means of presentation, narrowly construed, do not count all that much: one is obliged to pick out the right truth-conditions, up to logical equivalence. That is all that talk of 'formulae' need here achieve.

Basic Definitions and Explications

Some of the clauses in the definition are accompanied by short explanatory notes.

Definitions

1. For the sake of the following elucidatory definition, the base or 0-rank language $\mathcal{L}$ is assumed to be a second-order language in its ordinary, primitive recursive formulation. (In further instantiations of the definition, down the line, other base

languages will appear.) In addition to a countably infinite collection of predicate constants of assorted arities, $\mathcal{L}$ contains a countably infinite collection of individual variables denoted 'x' with or without subscripts, plus a disjoint countably infinite collection of predicate variables denoted 'X,' perhaps also with subscripts. In $\mathcal{L}$, there is a countably infinite collection of individual constants as well. **Note**: In general, $\mathcal{L}$ could be any formal or regimented natural language, and the correlative hierarchy of Tractarian W-formulae could be constructed over it with compensatory adjustments regarding variables bound in λ-abstractions.

2. In establishing notation for form-series, the definition of W-formula employs auxiliary bound individual variables y from a countably infinite collection fully disjoint from any collections of variables of $\mathcal{L}$.
3. A function from W-formulae to W-formulae is *primitive recursive* when it is primitive recursive as a function on the usual arithmetization or Gödel encoding of its inputs and outputs, that is, as a function from the natural numbers that encode its input W-formulae into those encoding its outputs. **Note**: All the syntactic mappings called 'operations' here and throughout are to be primitive recursive. The restriction to primitive recursive operations is natural in this context and consistent with Wittgenstein's several examples of operations, as at 4.1252, at 4.1273, and crucially at 6.02. Besides, Wittgenstein insists repeatedly that operations, associated with internal relations, 'show themselves' [*zeigt sich*] in the expressions of propositions; *vide* 4.122 and 5.24.
4. When $\phi(x)$ or $\phi(X)$ is a propositional function–of the sort soon to be defined–and x any suitable variable,

$$\lambda x \cdot \phi(x)$$

is a syntactic entity whose class is the collection of all legal substitution instances of the expression $\phi(x)$. **Note**: This notation is consonant with Wittgenstein's recommendations at 5.501 and elsewhere, and serves to replace the more familiar

$$\{\phi(y) : y \text{ is substitutable for } x \text{ in } \phi(x)\}.$$

The Russellian usage 'propositional function' does not feature as such in **Tractatus**; the adjective 'propositional' is employed to remind the reader that the values of a propositional function are strictly notational throughout.

5. Let ϕ be any W-formula, y an auxiliary variable whose range includes ϕ, and O an operation—a primitive recursive function as above. Then,

$$[\phi, y, O'y]$$

is the (*recursive*) *closure* of ϕ with respect to O, the least collection of W-formulae containing ϕ and closed under the operation mapping input values of y to output values $O'y$. The notation $[\phi, y, O'y]$ is Wittgenstein's. The denomination '(recursive) closure' is not. **Note**: For example–thinking of

conventional formulae–if ϕ is an individual formula and the operation is negation $\neg$, then $[\phi, y, \neg' y]$ is the least collection containing ϕ and closed under the prefixing of negations, provided that y is a variable containing ϕ in its class. Wittgenstein termed such linguistic formations 'bracket-expressions' [*Klammerausdruck*] for 'the general member of a form-series' [*Formenreihe*], rather than closures. See Remarks 4.1273 and 5.2522. Wittgenstein's examples make it plain that he was, in his own way, using bracket-expressions to demarcate (what we call) recursively defined collections.

6. When ξ is a (propositional) variable, then $\overline{\xi}$ denotes the class of all and only the values of ξ, as Wittgenstein explains at 5.501. **Note**: Incidentally, the use here of what are labeled 'classes' does not flout Wittgenstein's banishment of classes at 6.031,

   > In mathematics, the theory of classes is completely superfluous.

   The word is employed to convey to the reader, in the shortest intelligible form, the idea of the range of values of a variable that Wittgenstein endorses with the overline notation at 5.501. In the present article, all classes consist entirely of syntactical entities, in particular, of W-formulae of some rank. This usage of 'class' is in keeping with Wittgenstein's own: at 3.315 he employs the term 'class' [*Klasse*] to convey the same concept.
7. When C is a class of values of a propositional variable,

$$\bigwedge_{\phi \varepsilon C} \neg\phi$$

   denotes the (in general, infinitary) conjunction of the negations of all W-formulae ϕ that are members of the class C. **Note**: This notation is not Wittgenstein's, but is exploited to offer the reader a more accessible understanding of the scope and role of Wittgenstein's generalized Sheffer stroke or *N*-notation, as in Remarks 5.502ff. Once again, the reader is referred to the principle of extensionality.

Elements of the Notation

A reasonably accurate recapture of Wittgenstein's ideas, rendered via recursion, concerning the descriptions of variables and their roles in communicating propositions would seem to call for simultaneous delimitation of these six categories:

1. W-formulae,
2. operations,
3. lists, which Wittgenstein describes as 'direct enumeration' [*direkte Aufzählung*] at 5.501,

4. propositional functions,
5. form-series, and
6. classes.

Each finitary formula in any ordinary, primitive recursively defined formal language can be assigned a rank, a natural number marking the formula's complexity and its relative place in the definitional generation of the class of all formulae. In much the same spirit, each W-formula in the notation about to be defined belongs to one or another collection of n-formulae for some natural number n, the rank of the formula. The same goes for members of all the categories: lists, operations, and so on. A rank number 'n' marks the nth stage in the joint recursive definition; the first or lowest stage at which a formula appears is its proper rank. Although the W-formulae and members of other categories of notation are generated stepwise in the course of building the ω-series of collections of n-formulae, there is no type or order theory here: a rank n is not an index for a type or an order, as in **Principia Mathematica** (Whitehead and Russell 1950), that classifies semantically the items over which a variable can range or to which a formula can refer. The present ranks are, if you will, marks in a purely syntactical or formal classification. This is in keeping with the anti-type stricture Wittgenstein lays down at 3.332.

All stages in the recursion are divided into substages: those for W-formulae, operations, lists of formulae, propositional functions, form-series, and classes. The variables Wittgenstein requires explicitly at 5.501 are realized by the various substages for lists, propositional functions, and form-series. Substages of W-formulae are cumulative. For instance, W-formulae generated at Substage 1.0 are among those generated at Substage 2.0.

The specification at stage 0 is spelled out completely, as is the specification at stage $k + 1$; the general scheme for definition by recursion demands no less. Stage 1 appears for the sake of explanatory illustration only.

Definition: Stage 0

Once again, some of the substages are accompanied by short explanatory notes.

Substage 0.0 The collection of 0-*W-formulae* is the set of closed atomic formulae of $\mathcal{L}$. **Note**: These are stand-ins for Wittgenstein's elementary propositions [*Elementärsätze*] as introduced at 4.21. There is no suggestion that the 0-formulae exhibit all the logical and metaphysical properties that Wittgenstein expects of his elementary propositions, e.g., that they are no more than concatenations of unanalyzable names, as Remark 4.22 insists.

Substage 0.1 The collection of 0-*operations* is the set of all primitive recursive functions from 0-formulae into 0-formulae.

Substage 0.2 The collection of 0-*lists* is the set of all finite lists $\langle \phi, \psi, \chi \rangle$ of 0-formulae ϕ, ψ, and χ. **Note**: Here, as elsewhere, a three-place list serves as a paradigm only; lists of any finite, positive length may be formed.

Substage 0.3 The collection of 0-*propositional functions* is the set of closed λ-abstractions of 0-formulae with respect to one or more individual or second-order variables of $\mathcal{L}$. **Note**: For example, if $R(c, d)$ is a 0-formula with c and d individual constants of $\mathcal{L}$, then both $\lambda x \cdot [R(x, d)]$ and $\lambda x \lambda X \cdot [X(x, d)]$ are 0-propositional functions. These lambda expressions do not have extralinguistic functions as their denotations, but are variables ξ with classes $\overline{\xi}$ of substitution instances giving their ranges.

Substage 0.4 The collection of 0-*form-series* is the set of all expressions of the form

$$[\phi, y, O'y]$$

where ϕ is a 0-formula, O a 0-operation and y an auxiliary variable. **Note**: The idea is that '$[\phi, y, O'y]$' stands for the recursive closure, within the 0-formulae, of the singleton set $\{\phi\}$ with respect to the operation O. Please note that, here and elsewhere, every form-series is the class of some propositional variable and, as such, is itself a collection of W-formulae. Furthermore, it will be a collection of W-formulae naturally arranged by iterating the operation O:

$$x, O'(x), O'O'(x), O'O'O'(x), \ldots,$$

thus forming a strict ω-sequence. Consequently, the resultant class will be recursively enumerable.

The collection of 0-*variables* is the union of the previous three collections: lists, propositional functions, and form-series of rank 0, as 5.501 demands.

Substage 0.5 Closing off the 0-substage is collection of 0-*classes* containing all those sets obtained from a 0-variable ξ by passing to its overline $\overline{\xi}$, the symbol for its collection of values. **Note**: Expressed in more familiar terms, the collection of 0-classes circumscribes, for each 0-list, the set of its components; for each 0-propositional function, the set of all its legitimate substitution instances; and, for each 0-form-series, the set of all 0-formulae belonging to the respective recursive closure. Here, the legitimate substitution-instances of a 0-propositional function are restricted to W-formulae of rank 0.

Definition: Stage 1

Substage 1.0 The collection of 1-*W-formulae* is the set of all W-formulae constructed in the fashion Wittgenstein denoted

$$N(\overline{\xi}),$$

where ξ is a 0-variable, together with all 0-formulae. **Note**: Here, the W-formulae $N(\overline{\xi})$ are conceived as implemented, in the underlying logical operating system, if you will, by the more familiar forms

$$\bigwedge_{\phi \varepsilon C} \neg\phi,$$

where C is a 0-class. The latter is not Wittgenstein's notation but would be equivalent to it in contemporary eyes. More prosaically, one may think of '$N(\overline{\xi})$' as a helpful shorthand or for

$$\bigwedge_{\phi \varepsilon \overline{\xi}} \neg\phi,$$

for any 0-variable ξ.

Substage 1.1 The collection of 1-*operations* is the set of all primitive recursive functions from 1-W-formulae into 1-W-formulae. **Note**: The reader is reminded that the constraint that a 1-operation be primitive recursive is meaningful because, in virtue of their respective presentations as lists, propositional functions, or form-series, all 1-W-formulae possess Gödel codes that are standard natural numbers.

Substage 1.2 The collection of 1-*lists* is the set of all finite lists $\langle \phi, \psi, \chi \rangle$ of 1-W-formulae ϕ, ψ, and ξ.

Substage 1.3 The collection of 1-*propositional functions* is the set of closed λ-abstractions of 1-W-formulae with respect to one or more individual or second-order variables of $\mathcal{L}$. In this case, the substitution instances of a second-order λ-abstraction $\lambda X \cdot \phi$ of a 1-formula are obtained via admissible substitutions of 1-formulae. As usual, the instances of first-order abstractions are obtained by plugging individual constants of $\mathcal{L}$ into the W-formula that forms the body of the abstract.

Substage 1.4 The collection of 1-*form-series* is the set of all expressions of the form

$$[\phi, y, O'y],$$

where ϕ is a 1-W-formula, O a 1-operation, and y an auxiliary variable. **Note**: Once again, the idea is that '$[\phi, y, O'y]$,' indicates the recursive closure of the singleton set $\{\phi\}$ with respect to the function O within the collection of 1-W-formulae.

The collection of 1-*variables* is the union of the previous three collections: lists, propositional functions, and form-series of rank 1.

Substage 1.5 The collection of 1-*classes* contains all those sets obtained from a 1-variable ξ by passing to its overline $\overline{\xi}$, the symbol for its collection of values. **Note**: In more quotidian terms, the collection of 1-classes contains, for each 1-list, the set of its components; for each 1-propositional function, the set of all its

legitimate substitution instances; and, for each 1-form-series, the set of all 1-W-formulae $O^i(\phi) = O'O' \ldots O'(\phi)$ (with O repeated i times) for i a natural number, belonging to the respective recursive closure.
Now comes the general recursion step, the specification for stage $k+1$, with $k > 1$.

Definition: Stage k + 1

Substage $k+1.0$ The collection of $k+1$-*W-formulae* is the set of all W-formulae obtained in the fashion Wittgenstein denoted

$$N(\overline{\xi}),$$

where ξ is a k-variable, together with all k-W-formulae. **Note**: Just as before, the formulae $N(\overline{\xi})$ are here conceived as implemented in the form

$$\bigwedge_{\phi \varepsilon \overline{\xi}} \neg\phi,$$

where ξ is a k-variable.
Substage $k+1.1$ The collection of $k+1$-*operations* is the set of all primitive recursive functions from k-formulae into k-formulae. **Note**: Once again, by inductive reasoning, one checks that the constraint that a $k+1$-operation be primitive recursive remains meaningful because, through their presentations as lists, propositional functions, and form-series, all k-formulae possess codes.
Substage $k+1.2$ The collection of $k+1$-*lists* is the set of all finite lists $\langle \phi, \psi, \chi \rangle$ of $k+1$-W-formulae ϕ, ψ, and χ.
Substage $k+1.3$ The collection of $k+1$-*propositional functions* is the set of closed λ-abstractions of $k+1$-W-formulae with respect to one or more individual or second-order variables of $\mathcal{L}$. **Note**: As before, the substitution instances of a second-order λ-abstraction $\lambda X \cdot \phi$ of a $k+1$-formula, for example, must be obtained via admissible substitutions of $k+1$-W-formulae.
Substage $k+1.4$ The collection of $k+1$-*form-series* is the set of all expressions of the form

$$[\phi, y, O'y],$$

where ϕ is a $k+1$-W-formula, O a $k+1$-operation, and y an auxiliary. **Note**: '$[\phi, y, O'y]$' indicates the recursive closure of the singleton set $\{\phi\}$ with respect to the function O restricted to $k+1$-W-formulae.
The collection of $k+1$-*variables* is the union of the previous three collections: lists, propositional functions, and form-series of rank $k+1$.

Substage $k+1.5$ The collection of $k+1$-*classes* contains all those sets obtained from a $k+1$-variable ξ by passing to its overline $\overline{\xi}$, the symbol for its collection of values. **Note**: The collection of $k+1$-classes contains, for each $k+1$-list, the set of its components; for each $k+1$-propositional function, the set of all its legitimate substitution instances; and, for each $k+1$-form-series, the set of all $k+1$-W-formulae $O^i(\phi) = O'O' \cdots O'(\phi)$ (O repeated i times) for i a natural number.

Definition and Basic Properties of W-Formulae

Definitions The collection of *W-formulae* (in the hierarchy over $\mathcal{L}$) is the union of all the collections of n-W-formulae for all ranks n. Collections of *operations*, *propositional functions*, *lists*, *form-series*, and *classes* are definable *mutatis mutandis*.

Proposition *The ranking of W-formulae is cumulative: if n and m are ranks with* $n \leq m$, *every n-W-formula is also an m-W-formula.* ∎

With respect to the extension of the notion of W-formula, a major simplification in the definition of W-formulae is effected by noting the

Subsumption Lemma *At any rank k, the classes of W-formulae* $\overline{\xi}$ *generated as classes of lists and of propositional functions* ξ *can be formed from form-series variables alone without loss.*

Proof (by induction) At each rank, the classes derived from list variables and lambda abstraction variables are uniformly recursively enumerable. Therefore, those classes could have been presented as the recursive closures of suitable primitive recursive operations. ∎

For the sake of exposition and to maintain close connection with Wittgenstein's specification of 5.501, reference to variables given by lists and by propositional functions, although strictly unnecessary from an extensional viewpoint, will continue to feature in occasional examples and proofs.

Conventional Propositional Formulae and Expressive Completeness

It is readily seen that a suitable construction along the lines of the above, but with a standard language for ordinary, finitary propositional logic as base, incorporates the expressive power of all formulae of a language of conventional, finitary propositional logic.

For purposes of the current section, let $\mathcal{L}$ in the above schematic specification of the hierarchy of W-formulae be replaced by a standard language for propositional logic $\mathcal{L}^{prop}$ with a countably infinite set **At** of atomic formulae

$$p_0, p_1, p_2, \ldots$$

and with the ordinary, binary Sheffer stroke | as sole connective. In this case, the collection of propositional functions, substage $n + 1.3$ of each rank $n + 1$ in the definition of the hierarchy, will be considered empty. This decision does not affect the (extensional) set of W-formulae that the resultant definition generates. See the Subsumption Lemma of the previous section.

Definitions

1. **T** and **F** are the members of the familiar, two-element, complete Boolean algebra $\mathfrak{B}$ of truth-values with complementation operation $\neg$, greatest lower bound Λ, and least upper bound $\vee$.
2. Any function from **At** into $\mathfrak{B}$ is an *assignment*. $\mathfrak{A}$ is the set of all assignments.
3. Any function from $\mathfrak{A}$ into $\mathfrak{B}$ is an *infinitary truth-table*.
4. For any natural number n, a function from $\mathfrak{B}^n$, the n-fold Cartesian power of $\mathfrak{B}$, into $\mathfrak{B}$ is a *finitary truth-table* of arity n.
5. Pairs of ordinary, propositional formulae and W-formulae are *logically equivalent in propositional logic* just in case they share an infinitary truth-table.

Proposition *Let a be any assignment. There is a unique natural interpretation function*

$$[[\ldots]]_a : \phi \rightarrow [[\phi]]_a,$$

mapping each W-formula ϕ into $[[\phi]]_a \varepsilon\, \mathfrak{B}$, commuting with the connectives, and extending a.

Proof (definition by recursion and proof by induction on ranks n) For W-formulae ϕ of rank 0, i.e., formulae $p_0, p_1, p_2, \ldots$,

$$[[\phi]]_a = a(\phi).$$

If $[[\phi]]_a$ is defined for all W-formulae of rank n, then a W-formula of rank $n + 1$ is

$$N(\overline{\xi})$$

where ξ is either a list or a form-series of W-formulae of rank n. In the first case, if

$$\xi = \langle \phi, \psi, \chi \rangle,$$

then

$$[[N(\overline{\xi})]]_a = \Lambda\{\neg[[\phi]]_a, \neg[[\psi]]_a, \neg[[\chi]]_a\}.$$

In the second case,

$$\xi = [\phi, y, O'y],$$

where ϕ is a W-formula of rank n and O is an n-operation. Let C be the class of ξ, that is, $\overline{\xi}$. Then

$$[[N(\overline{\xi})]]_a = \bigwedge_{\phi \varepsilon C} \neg [[\phi]]_a.$$

Since $\mathfrak{B}$ is a complete Boolean algebra, one employs the glb operation in $\mathfrak{B}$ to give the definition of $[[\ldots]]_a$ in each of these cases. ∎

Corollary *Every* W-formula ϕ *has a unique infinitary truth-table given by the map* $\lambda a \varepsilon \mathfrak{A} \cdot [[\phi]]_a : \mathfrak{A} \rightarrow \mathfrak{B}$. ∎

Familiar finitary formulae in the language $\mathcal{L}^{prop}$ are also associated with unique infinitary truth-tables in the usual fashion. Moreover, each W-formula containing at most a finite number of atomic formulae of $\mathcal{L}^{prop}$ can be attached to a unique finitary truth-table, too.

It is obvious that the W-formulae do not constitute an expressively complete collection with respect to infinitary truth-tables: there are only $\aleph_0$-many W-formulae but $\beth_2$-many infinitary truth-tables. However, the W-formulae clearly suffice to express every finitary truth-table:

Theorem *Given any formula* ψ *of* $\mathcal{L}^{prop}$*, there is a W-formula* ϕ *such that* ϕ *and* ψ *are logically equivalent in propositional logic.*

Proof (by induction on formulae ψ of $\mathcal{L}^{prop}$) Every atomic formula p of $\mathcal{L}^{prop}$ is already a 0-W-formula. Assume that formulae ψ_1 and ψ_2 of $\mathcal{L}^{prop}$ have the same infinitary truth-tables as W-formulae ϕ_1 and ϕ_1, respectively. Because the ranks in the hierarchy of W-formulae are cumulative, there is a rank n to which both ϕ_1 and ϕ_2 belong. Then,

$$\xi = \langle \phi_1, \phi_2 \rangle$$

is a finite list of n-W-formulae, and $N(\overline{\xi}) = \text{⫫}\{\neg\phi_1$, is an $n+1$-W-formula. Plainly, $\psi_1|\psi_2$ and $N(\overline{\xi})$ share an infinitary truth-table. ∎

Corollary *For every finitary truth-table, there is a W-formula in the hierarchy that has that truth-table.*

Proof The set $\{|\}$ is an expressively complete set of connectives. ∎

Proposition *The collection of 4-W-formulae is expressively complete with respect to finitary truth-tables.*

Proof A quick calculation with finitary disjunctive normal forms suffices. ∎

Classical Second-Order Propositional Logic

The W-formulae are sufficient to the expression of the semantical values of formulae in classical second-order propositional logic.

Once again, let $\mathcal{L}^{prop}$ be a standard language of propositional logic with a countably infinite set **At** of atomic formulae

$$p_0, p_1, p_2, \ldots$$

and with Sheffer stroke | as sole connective. Also as before, the collection of propositional functions, substage $n+1.3$ of each rank $n+1$ in the recursive definition of the hierarchy, will be empty.

Definitions

1. Let $\mathcal{L}^2$ be the obvious, minimal, second-order extension of $\mathcal{L}^{prop}$, obtained by adjoining to the latter language a countably infinite collection of quantifiable propositional variables $v_0, v_1, v_2, \ldots$ as atomic symbols, so expanding the set **At**, and then closing the resultant collection of formulae under the prefixing of quantifiers $\forall v$ and $\exists v$, where v is a propositional variable.
2. Any function from **At** into $\mathfrak{B}$ is an *assignment*. Once again, $\mathfrak{A}$ is the set of all assignments.
3. Any function from $\mathfrak{A}$ into $\mathfrak{B}$ is an *infinitary truth-table*.
4. For any assignment a, propositional variable v, and $b \varepsilon \mathfrak{B}$, the assignment $a[v \rightarrow b]$, the *b-variant of a at v*, is identical to a except, perhaps, that v maps to the truth-value b.
5. Let T and F be two fixed formulae of $\mathcal{L}^2$, the first a standard tautology and the second a standard contradiction.
6. For ψ a formula of $\mathcal{L}^2$, $\psi[v \rightarrow \mathrm{T}]$ is the result of substituting T for all free appearances of propositional variable v in ψ, and similarly for $\psi[v \rightarrow \mathrm{F}]$.
7. Pairs of ordinary, second-order formulae and W-formulae are *logically equivalent in second-order propositional logic* just in case the formulae share the same infinitary truth-table.

Again, one erects the hierarchy of W-formulae over the atomic formulae of $\mathcal{L}^{prop}$. Since $\mathcal{L}^{prop}$ features no individual first- or higher-order variables of quantification, the collection of propositional functions, substage $n+1.3$ of each rank $n+1$ in the definition of the hierarchy, is assumed to be empty. This choice does not affect the extension of the defined notion 'W-formula.'

Proposition *Let a be an assignment of truth-values **T** and **F** to all atomic formulae of* $\mathcal{L}^2$. *There is a unique natural interpretation function*

$$[\![\ldots]\!]_a : \phi \rightarrow [\![\phi]\!]_a,$$

mapping each formula ϕ *of* $\mathcal{L}^2$ *into* $[[\phi]]_a \varepsilon \mathfrak{B}$ *and extending a.*

Proof (definition by recursion and proof by induction on formulae of $\mathcal{L}^2$) Extend the usual notion of interpretation for strictly propositional $\mathcal{L}^{prop}$ with the specifications

$$[[\forall v \cdot \phi]]_a = \bigwedge_{b \varepsilon \mathfrak{B}} [[\phi]]_{a[v \to b]}, \quad \text{and}$$

$$[[\exists v \cdot \phi]]_a = \bigvee_{b \varepsilon \mathfrak{B}} [[\phi]]_{a[v \to b]}. \quad \blacksquare$$

Corollary *Each* formula ϕ *in* $\mathcal{L}^2$ *has a unique infinitary truth-table given by the map* $\lambda a \varepsilon \mathfrak{A} \cdot [[\phi]]_a : \mathfrak{A} \to \mathfrak{B}$. ■

Since the Sheffer stroke is expressively complete as a propositional connective, there are standard representations in $\mathcal{L}^2$ for propositional conjunction $\wedge$ and disjunction $\vee$.

Theorem *Given any formula* ψ *of* $\mathcal{L}^2$*, there is a W-formula* ϕ *such that* ϕ *and* ψ *are logically equivalent in second-order propositional logic.*

Proof Quantifiers are eliminable from classical second-order propositional logic under its standard semantics. For instance, $\forall v \cdot \chi$ has the same infinitary truth-table as the standard formula

$$\chi[v \to \mathrm{T}] \wedge \chi[v \to \mathrm{F}].$$

In the same spirit, $\exists v \cdot \chi$ has the same infinitary truth-table as

$$\chi[v \to \mathrm{T}] \vee \chi[v \to \mathrm{F}].$$

The theorem now follows from the results of the preceding section. ■

Arithmetic and Arithmetic Sets

This time, the hierarchy of Tractarian W-formulae is defined over a standard language for arithmetic, $\mathcal{L}^A$, and one proves that the sets of and relations on numbers definable by W-formulae in parameters are precisely Kleene's arithmetic sets and relations (Rogers 1967, 301ff).

Definitions

1. $\mathcal{L}^A$ is a standard, first-order formal language for elementary Peano-Dedekind arithmetic in which each n-ary primitive recursive relation on the natural numbers is represented specifically by a distinct atomic formula

$$A(x_0, x_1, \ldots, x_{n-1})$$

of the language. It is assumed that the atomic (and, hence, arbitrary) formulae of $\mathcal{L}^A$ feature individual, numerical parameters subject neither to binding within $\mathcal{L}^A$ nor to λ-abstraction within the construction of the hierarchy of W-formulae over the base $\mathcal{L}^A$.

2. In $\mathcal{L}^A$, $\underline{m}$ is the conventional numeral denoting natural number m.
3. Let $\mathfrak{N}$ be the standard interpretation for the language $\mathcal{L}^A$.
4. An assignment a is a function from the set of all variables and parameters of $\mathcal{L}^A$ into the domain of interpretation $\mathfrak{N}$. For a an assignment, $\vec{x}$ an n-tuple of parameters, and $\vec{m}$ an n-tuple of natural numbers, then $a[\vec{x} \rightarrow \vec{m}]$ is an assignment identical to a except in attaching, to each component of $\vec{x}$, the corresponding component of $\vec{m}$.
5. The various labels 'Σ_k' and 'Π_k' for levels in the Kleene hierarchy are here employed not only to name those levels but also to indicate the correlative quantifier-complexity of $\mathcal{L}^A$ formulae. For example, a set of numbers is Σ_2 when it is definable over $\mathfrak{N}$ by a formula that is also Σ_2. Context should suffice to disambiguate.

Satisfaction for W-formulae with respect to $\mathfrak{N}$ is definable in the expected recursive fashion. For example, let ξ be an n-variable that is a finite list of components ϕ, and let a be an assignment. In that case,

$$\mathfrak{N} \vDash N(\overline{\xi})[a] \text{ iff}, \quad \text{for all components } \phi \text{ in list } \xi, \quad \mathfrak{N} \nvDash \phi[a].$$

If ξ is an n-variable that is a form-series $[\phi, y, O, y]$ and a an assignment, then

$$\mathfrak{N} \vDash N(\overline{\xi})[a] \text{ iff}, \quad \text{for all natural numbers } i, \quad \mathfrak{N} \nvDash O^i(\phi)[a].$$

More Definitions:

1. An n-ary relation R on the natural numbers is *W-definable* just in case there is a W-formula ϕ in n distinct parameters in the hierarchy over $\mathcal{L}^A$ such that, for any assignment a and for any n-tuple $\vec{m}$ of numbers, $\vec{m} \varepsilon R$ if and only if $\mathfrak{N} \vDash \phi[a[\vec{x} \rightarrow \vec{m}]]$. The W-formula ϕ is then said to *W-define* the relation R.
2. A pair of $\mathcal{L}^A$ formulae and W-formulae are *arithmetically equivalent* just in case they define or W-define the same set or relation of numbers.

Theorem *Every arithmetical set and relation of numbers is W-definable.*

Proof (by induction on the levels of Kleene's hierarchy) By definition of $\mathcal{L}^A$, every primitive recursive set of numbers is definable by a 0-W-formula.

Since W-definability is obviously closed under complementation with respect to the natural numbers, it suffices to assume that all Σ_k and Π_k sets and relations are W-definable, and, on that assumption, to prove that all Π_{k+1} sets are W-definable. Let $S \varepsilon \Pi_{k+1}$. Then, there exists a relation $T \varepsilon \Sigma_k$ such that, for all n,

$$n \varepsilon S \text{ if and only if } \mathfrak{N} \vDash \forall m \cdot T(\underline{n}, m).$$

Hence, for all n, informally,

$$n \varepsilon S \text{ if and only if } \mathfrak{N} \vDash T(\underline{n}, \underline{0}) \wedge T(\underline{n}, \underline{1}) \wedge \ldots.$$

By assumption, there is a W-formula ϕ with two parameters such that ϕ W-defines T. From the construction of the hierarchy, one knows that, if $\langle \phi \rangle$ is a one-place list, then

$$\psi = N\left(\overline{\langle \phi \rangle}\right)$$

is a W-formula as well. Let O be the operation that maps $\psi(\underline{n}, \underline{0})$ into $\psi(\underline{n}, \underline{1})$ and, in general, $\psi(\underline{n}, \underline{k})$ into $\psi(\underline{n}, \underline{k+1})$. In consequence,

$$[\psi(\underline{n}, \underline{0}), y, O'y]$$

is a form-series, and

$$N(\overline{[\psi(\underline{n}, \underline{0}), y, O'y]})$$

is a W-formula. It is easy to see that $N(\overline{[\psi(\underline{n}, \underline{0}), y, O'y]})$ W-defines S. ∎

The converse of the above theorem holds, too; every W-definable set or relation of numbers is arithmetic.

Theorem *Every W-definable set of numbers is arithmetic.*

Proof (by induction on ranks of W-formulae) What we here prove is the stronger assertion that there is a primitive recursive function *tr* mapping the (codes of) W-formulae to (the codes of) formulae of $\mathcal{L}^A$ such that ϕ and $tr(\phi)$ define the same set of (or relation on) numbers, and *tr* maps the n-W-formulae uniformly into arithmetic formulae in $\Sigma_l \cup \Pi_l$, for some fixed l depending upon n. We exploit the fact that a truth definition for formulae naturally defining sets and relations in $\Sigma_k \cup \Pi_k$ is contained in Π_{k+1}.

By definition of $\mathcal{L}^A$, every W-formula of rank 0 defines a primitive recursive set of numbers. In that case, *tr* is the identity on W-formulae of rank 0 and maps those formulae into $\Sigma_0 \cap \Pi_0$.

Assume that all W-formulae of rank k define arithmetic sets or relations, and that *tr* works on k-W-formulae as specified above, mapping them into some fixed level $\Sigma_l \cap \Pi_l$. Variables ξ of rank k are either (i) finite lists of k-W-formulae, (ii) propositional functions obtained from k-W-formulae, or (iii) form-series defined over k-W-formulae. A W-formula of rank $k+1$ is formed by prefixing the N operator to the class of a variable of rank k to obtain $N(\overline{\xi})$. In case (i), ξ is the list

$$\langle \phi, \psi, \chi \rangle$$

where ϕ, ψ, and χ are k-W-formulae, and $N(\overline{\xi})$ is 'really' the formula

$$\neg\phi \wedge \neg\psi \wedge \neg\chi.$$

By assumption, $tr(\phi)$, $tr(\psi)$, and $tr(\chi)$ are formulae in $\Sigma_l \cup \Pi_l$ that define the same sets or relations as their W-counterparts, respectively. So, the arithmetic formula

$$\neg tr(\phi) \wedge \neg tr(\psi) \wedge \neg tr(\chi)$$

defines the same set or relation as $N(\overline{\xi})$ and can be assumed to be at level $\Sigma_l \cup \Pi_l$. Hence, *tr* can be extended recursively and uniformly to $k+1$-W-formulae constructed from list variables of rank k.

In case (ii), ξ is the propositional function $\lambda x \cdot \phi$ where ϕ is a k-W-formula. $N(\overline{\xi})$ is therefore the $k+1$-formula expressed informally as

$$\neg\phi(\underline{0}) \wedge \neg\phi(\underline{1}) \wedge \ldots.$$

Plainly, all the W-formulae $\phi(\underline{n})$ for n a natural number are also k-W-formulae, as a simple inductive argument confirms. By assumption, there are arithmetic formulae $tr(\phi(\underline{n}))$ defining the same set or relation as $\phi(\underline{n})$, respectively, and lying within $\Sigma_l \cup \Pi_l$ for some l. Because truth for arithmetic formulae in $\Sigma_l \cup \Pi_l$ is definable in $\Sigma_{l+1} \cup \Pi_{l+1}$, there is an arithmetic formula ψ in $\Sigma_{l+1} \cup \Pi_{l+1}$ such that, for all assignments a, $\mathfrak{N} \vDash \psi[a]$ if and only if

$$\forall i \; \mathfrak{N} \nvDash tr(\phi(\underline{i}))[a].$$

Therefore, ψ defines the same set or relation as $N(\overline{\xi})$, and *tr* can be extended primitive recursively and uniformly to $k+1$-W-formulae constructed from propositional functions of rank k.

The argument for case (iii) runs parallel to that for (ii). ∎

Because of the treatment of variables ξ determined by propositional functions in the definition of the hierarchy of W-formulae, the previous theorems continue to hold when $\mathcal{L}^A$ is extended to allow second-order quantification.

Noncollapsing Theorems

The hierarchies of W-formulae over $\mathcal{L}^A$ and over $\mathcal{L}^{prop}$ do not collapse.

Arithmetic Noncollapsing Theorem *For any rank n in the hierarchy of W-formulae constructed over* $\mathcal{L}^A$, *there is a W-formula of rank greater than n that is not arithmetically equivalent to any n-W-formula.*

Proof This follows immediately from the proofs of the preceding two theorems. ∎

Propositional Noncollapsing Theorem *For any rank n in the hierarchy of W-formulae constructed over* $\mathcal{L}^{prop}$, *there is a W-formula of rank greater than n that is not logically equivalent in propositional logic to any n-W-formula.*

Proof (by induction on rank) If two W-formulae in the hierarchy over $\mathcal{L}^{prop}$ are logically equivalent in propositional logic, then the results of uniform substitutions of atomic formulae of $\mathcal{L}^A$ for atomic formulae of $\mathcal{L}^{prop}$ would be W-formulae of the hierarchy over $\mathcal{L}^A$ that are arithmetically equivalent. ∎

Disjunctive Normal Forms and Standard Representations

Because every W-formula in the hierarchy over $\mathcal{L}^{prop}$ has an infinitary truth-table, it is logically equivalent in propositional logic both to a conventional infinitary formula in disjunctive normal form, as well as to one in conjunctive normal form. However, it is false that, in every case, these normal forms are themselves W-formulae. In extensional effect, all of the latter formulae are well-founded trees of finite height with recursively enumerable branching at every node.

Definitions

1. A *literal* is either a 0-W-formula or the negation of such a W-formula. For W-formulae p, the negation of p is $N\left(\overline{\langle p \rangle}\right)$.
2. A W-formula is a *disjunctive normal form* (*DNF*) just in case it is, in effect, a disjunction of conjunctions of literals, where those disjunctions and conjunctions may be infinitary.

It is easy to represent disjunction, finitary or recursively enumerable, in canonical terms using Wittgenstein's N operator and his notion of variable. In the finite case, $\phi \vee \psi$ is equivalent to $N\left(\left\langle N\overline{\langle \phi, \psi \rangle}\right\rangle\right)$. Hence, it is meaningful to speak unambiguously of 'disjunctions' in the case of W-formulae over $\mathcal{L}^{prop}$. More formally,

More Definitions

1. The function that takes a finite (or infinite) list $\langle \phi, \psi \rangle$ of W-formulae in the hierarchy over $\mathcal{L}^{prop}$, all of bounded rank, into the W-formula

$$N\left(\overline{\left\langle N\overline{\langle \phi, \psi \rangle}\right\rangle}\right)$$

   is represented by $\bigvee\!\!\bigvee$.
2. A W-formula in the hierarchy over $\mathcal{L}^{prop}$ is a (or is in) *W-disjunctive normal form* or *W-DNF* just in case it is of the form

$$\bigvee\!\!\bigvee_{i \varepsilon D} \bigwedge\!\!\bigwedge_{\phi \varepsilon C_i} \neg\phi,$$

   wherein each ϕ is a literal. D is a set of natural numbers.

Proposition Every W-formula over $\mathcal{L}^{prop}$ that is in W-DNF has rank at most 4.

Proof A simple calculation on 0-W-formulae, plus variables that are lists or form-series, suffices. ∎

Theorem *There are W-formulae in the hierarchy over $\mathcal{L}^{prop}$ that are not logically equivalent in propositional logic to any W-formula in the same hierarchy that is in W-DNF.*

Proof Follows immediately from the Propositional Noncollapsing Theorem of the preceding section. ∎

One can define a set of more familiar formulae, C-formulae, that capture the exact logical strength of the W-formulae over $\mathcal{L}^{prop}$. Here, one is once again working the qualification that Wittgenstein set down in Remark 5.501: 'How the description of the components of the bracket-expression comes about is inessential.' As was asserted, this is a principle of extensionality, informing readers that the class of propositions (formulae, in our terms) captured by the W-formulae is the very same, logically speaking, as that captured by the C-formulae defined below. At this point, one should recall that, for Wittgenstein, two logically equivalent sentential expressions yield the very same proposition. See Remarks 4.4 and 4.431.

Definitions

1. Let $\bigvee\!\!\bigvee_{re}$ and $\bigwedge\!\!\bigwedge_{re}$ represent the functions yielding disjunctions, respectively conjunctions, of recursively enumerable collections of formulae.
2. Let the collection of *C-formulae* (over $\mathcal{L}^{prop}$) be the least set containing the atomic formulae of $\mathcal{L}^{prop}$ and closed under the formula-building functions $\neg$, $\bigvee\!\!\bigvee_{re}$, and $\bigwedge\!\!\bigwedge_{re}$.

Just as with W-formulae, one can conceive of the C-formulae as decorated, well-founded trees organized into finite *levels*. On one way of viewing things, the C-formulae are those infinitary formulae whose canonical parsing trees are recursive ordinals à la Church and Kleene (Rogers 1967, 205ff).

Theorem *Every* W-formula *in the hierarchy over* $\mathcal{L}^{prop}$ *is logically equivalent in propositional logic to a C-formula, and conversely.*

Proof The left-to-right direction is obvious, given the Subsumption Lemma and that every W-formula deriving from a form-series is just another label for a C-formula.

The converse direction proceeds by induction on levels, proving the statement (1) that there is a primitive recursive function *tr* that converts any C-formula into a logically equivalent W-formula, and (2) that *tr* converts C-formulae of a single level into W-formulae of a single rank.

In the case of atomic formulae of $\mathcal{L}^{prop}$, the statements (1) and (2) are immediate.

Assume that *tr* converts the C-formulae ϕ_i of level n, where the i's are members of a recursively enumerable set S of numbers, into W-formulae of rank m so that each ϕ_i is logically equivalent to $tr(\phi_i)$. Then, $\neg\phi_i$, for any i, is equivalent to the W-formula $N\left(\overline{\langle tr(\phi_i)\rangle}\right)$, which has rank $m+1$.

Let f be a primitive recursive natural number function enumerating S. Let O be a primitive recursive operation mapping $tr(\neg\phi_{f(n)})$ into $tr(\neg\phi_{f(n+1)})$ for all n. Then, when $\xi = [tr(\neg\phi_{f(0)}), y, O^{\cdot}y]$, $\bigwedge\!\!\!\bigwedge_{i \varepsilon S} \phi_i$ is logically equivalent to $N(\overline{\xi})$, which has rank $m+2$.

The reasoning in the case of $\bigvee\!\!\!\bigvee_{re}$ is analogous. ∎

Expressibility and Inexpressibility in Predicate Logic

Let $\mathcal{L}^{pred}$ be any standard formal language for first-order predicate logic with a countable infinity of individual constants, plus a countable infinity of predicate constants. Let ψ be any atomic formula of $\mathcal{L}^{pred}$ in one free variable x. There is no W-formula in the hierarchy over $\mathcal{L}^{pred}$ that is logically equivalent to either $\forall x\psi$ or $\exists x\psi$, as logical equivalence is now commonly conceived.

Definitions

1. $\mathcal{L}^{pred}$ is a standard formal language for first-order predicate logic featuring a countable infinity of distinct individual constants c_i for i a natural number, and a countable infinity of distinct predicate constants of each arity.
2. $\mathfrak{A}$ is an arbitrary Tarskian *interpretation* for $\mathcal{L}^{pred}$ and a is an *assignment* function mapping the variables of $\mathcal{L}^{pred}$ into the domain of $\mathfrak{A}$.

3. *Satisfaction* $\vDash$ with respect to $\mathfrak{A}$ and assignment a are defined for formulae of $\mathcal{L}^{pred}$ in the canonical fashion and for W-formulae in the hierarchy constructed over $\mathcal{L}^{pred}$ in the expected fashion, as exemplified above.
4. W-formula ϕ in the hierarchy over $\mathcal{L}^{pred}$ and formula ψ of $\mathcal{L}^{pred}$ itself are *logically equivalent in predicate logic* if and only if, for all interpretations $\mathfrak{A}$ and assignments a,

$$\mathfrak{A} \vDash \phi[a] \text{ if and only if } \mathfrak{A} \vDash \psi[a].$$

5. An interpretation $\mathfrak{A}$ of $\mathcal{L}^{pred}$ is *full* if and only if every member of the domain of $\mathfrak{A}$ is the denotation of some individual constant of $\mathcal{L}^{pred}$.
6. W-formula ϕ in the hierarchy over $\mathcal{L}^{pred}$ and formula ψ of $\mathcal{L}^{pred}$ itself are 2-*logically equivalent* if and only if, for all full interpretations $\mathfrak{A}$ and assignments a,

$$\mathfrak{A} \vDash \phi[a] \text{ if and only if } \mathfrak{A} \vDash \psi[a].$$

Theorem *Let $P(x)$ be a unary predicate constant of $\mathcal{L}^{pred}$. There is no W-formula in the hierarchy over $\mathcal{L}^{pred}$ that is logically equivalent in predicate logic to either $\exists x \cdot P(x)$ or $\forall x \cdot P(x)$.*

Proof Since W-formulae are closed under negation, $\exists x \cdot P(x)$ is logically equivalent in predicate logic to a W-formula just in case $\forall x \cdot P(x)$ is. Assume that ϕ in the W-hierarchy over $\mathcal{L}^{pred}$ is equivalent in predicate logic to $\exists x \cdot P(x)$.

Let $\mathfrak{A}$ be a structure whose domain is the singleton set $\{0\}$, and let the denotation of P in $\mathfrak{A}$ be the empty set $\emptyset$. Assign $\mathfrak{A}$-denotations to all other predicate and relation constants of the formal language so that the resultant denotations hold of 0 or n-tuples of 0, as required. All individual constants denote 0 in $\mathfrak{A}$. Let a be the assignment that assigns 0 to all individual variables. Then $\mathfrak{A} \nvDash \exists x \cdot P(x)[a]$. By the assumption of equivalence, one concludes that $\mathfrak{A} \nvDash \phi[a]$.

Let $\mathfrak{B}$ be the interpretation that is just like $\mathfrak{A}$ except that the domain of $\mathfrak{B}$ is $\{0, 1\}$, and the denotation of P in $\mathfrak{B}$ is $\{1\}$. Let b be a $\mathfrak{B}$-assignment that is the same as a. Since the $\mathfrak{B}$ interpretations of linguistic elements of ϕ do not differ from their $\mathfrak{A}$ interpretations, $\mathfrak{B} \nvDash \phi[b]$. However, it is obvious that $\mathfrak{B} \vDash \exists x \cdot P(x)$. ∎

Theorem *Every formula of $\mathcal{L}^{pred}$ is 2-logically equivalent to some W-formulae in the hierarchy over $\mathcal{L}^{pred}$.*

Proof (by induction on the structure of formulae in $\mathcal{L}^{pred}$) From the definition of W-formula, it follows that every atomic formula of $\mathcal{L}^{pred}$ is 2-logically equivalent to a W-formula. Now, assume that $\phi(x, y)$ is 2-logically equivalent to some W-formula $\psi(x, y)$. One shows that $\forall x \cdot \phi(x, y)$ is also 2-logically equivalent to some W-formula. Remembering that the collection of W-formulae is closed under negation, and constructing the relevant form-series, one sees that the expression

$$\bigwedge_i \neg\neg\psi(c_i, y)$$

is (2-logically equivalent) to a W-formula that is 2-equivalent to $\forall x \cdot \phi(x, y)$. ∎

The proof of the preceding inexpressibility theorem would have been available, in principle, to Wittgenstein in 1921 since, to prove that result, no serious model theory is required. It would have been sufficient to note, for example, that the real numbers form an extension of the rational numbers in which not all members get named. The simple construction of alternative interpretations $\mathfrak{A}$ and $\mathfrak{B}$ could well have been represented as truth-table rows in a diagram such as that at 4.442. However, because the simple objects of **Tractatus** are denoted by Tractarian names, the structures corresponding to Wittgenstein's truth-table rows would have to be full. Hence, one must say that Wittgenstein's own logical theory in **Tractatus**, in which all interpretational structures are in effect full, could not itself have supported the argument to prove that $\forall x \cdot P(x)$ is not expressible by W-formulae.

Closing the Circle: Remark 6 Reconstructed

As asserted earlier, as far as the logic of Wittgenstein's **Tractatus** is concerned, the most prominent remark may be that numbered 6:

> The general form of the truth-function is: $[\overline{p}, \overline{\xi}, N(\overline{\xi})]$.
> This is the general form of the proposition.

On the present reconstruction, the natural rendering of the class of values for the expression '$[\overline{p}, \overline{\xi}, N(\overline{\xi})]$' is the *entire* form-series or recursive definition that defines the serial hierarchy of W-formulae itself: the collections of 0-formulae followed by that of 1-formulae, followed by that of 2-formulae, and so on. The collection of 0-formulae is just the class of values of the variable $\overline{p}$, namely the collection of closed atomic formulae of $\mathcal{L}$. As our recursive specification shows, one passes, in general, from propositional formulae of one rank to propositional formulae of the successor rank by combining a class of formulae of the former rank with the N operator. The class of the propositional variable ξ is a class C of propositional formulae of some rank n, and the class of $N(\overline{\xi})$ is the collection of results of applying the joint denial operation N, or the general operation taking class C into

$$\bigwedge\!\!\!\bigwedge_{\phi \varepsilon C} \neg\phi,$$

yielding W-formulae of rank $n + 1$. Wittgenstein does permit–and explicitly–such large-scale operations. See Remark 5.252.

Finally, there is some evidence, e.g., at (Wittgenstein 1980, 119), that Wittgenstein may indeed have conceived of the entire hierarchy $[\overline{p}, \overline{\xi}, N(\overline{\xi})]$ as itself

forming a new and ∞th rank in the recursive generation. However, we will not here pursue that suggestion further.

Acknowledgments A lecture based on this article was delivered to the seminar of the Indiana University Logic Program. The authors are grateful to Professor Larry Moss, seminar organizer, and to the participants of the seminar for their comments, questions, and suggestions. Special mention goes to Professor Gary Ebbs for inquiring if the semantical methods employed in proving the inexpressibility of quantified formulae of predicate logic would have been available, even in principle, to Wittgenstein at the time he wrote **Tractatus**.

References

Beth, E. W. (1955). *Semantic entailment and formal derivability*. Mededelingen van de Koninklijke Nederlandse Akademie van Wetenschappen. Afdeling Letterkunde. N.R. (Vol. 18, No. 13, pp. 309–342).

Dedekind, R. (1888) *Was Sind und was Sollen die Zahlen?* (xviii+58) Braunschweig, DE: Vieweg und Sohn.

Frege, G. (1893/1903). *Grundgesetze der Arithmetik. Bände I/II* (xxxii+253 (I), xv+265 (II)). Jena, DE: Verlag von Hermann Pohle.

Hintikka, M., & Hintikka, J. (1986). *Investigating Wittgenstein* (xii+248). Oxford, UK: Basil Blackwell.

Rogers, H. (1967). *Theory of Recursive Functions and Effective Computability* (xix+482). New York, NY: McGraw-Hill Book Company.

Skolem, T. (1923). *Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichen Ausdehnungsbereich*. Skrifter. Norske Videnskaps-Akademi i Oslo (Vol. I, No. 6b, pp. 1–38).

Whitehead, A. N., & Russell, B. (1950). *Principia Mathematica* (Vols. I, II, III, 2nd Edn, xlvi+674 (I), xxxi+742 (II), viii+491 (III)). Cambridge, UK: Cambridge University Press.

Wittgenstein, L. W. (1980). *Wittgenstein's lectures* (xvii+124). Cambridge, 1930–1932. (From the Notes of J. King & D. Lee (Ed.)). Chicago, IL: University of Chicago Press.

Wittgenstein, L. W. (1984). Tractatus Logico-Philosophicus. Werkausgabe Band 1 (611 pp.). Frankfurt, DE: Suhrkamp Verlag (References by remark number).

Justifying Knowledge Claims After the Private Language Argument

Gheorghe Ştefanov

I am trying, in what follows, to show, by way of an example, that Wittgenstein's thinking can still foster some interesting contributions to philosophy. The example consists of my own attempt to propose a way in which we could better conceive the justification of our knowledge claims. First, I offer my understanding of the Private Language Argument (PLA) and of the challenges it raises for the empiricist foundationalist. Second, I argue that one may find some interesting suggestions for a better way to conceive knowledge justification in Wittgenstein's *On Certainty*. At the end of my chapter I consider some possible objections to my conceptual proposal.

In short, I take the PLA to offer support to the claim that no experience, conceived as an inner episode to which only the subject having it has direct access, can be semantically relevant. A direct consequence of this would be that no experience thus conceived can be epistemically relevant. If this is accepted, then the traditional empiricist project fails. Some newer takes on the project, however, appear to be unaffected. This requires closer scrutiny, but let us first see how the initial claim is given support in *Philosophical Investigations*.

I start with a reconstruction of Wittgenstein's arguments involving one of the messages derived from the Rule Following Considerations, namely that every concept must presuppose a practice. Pace Kripke,[1] I do not think that the idea that it is impossible for an isolated subject to name his or her sensations follows immediately from the previous assumption, because one could invent a practice, as Wittgenstein himself does quite often. In fact, the remarks grouped under the PLA tag start with the observation that there are practices which a human being could be involved in alone:

[1]See Kripke (1982).

G. Ştefanov (✉)
University of Bucharest, Bucharest, Romania
e-mail: gstefanov@gmail.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_13

> A human being can encourage himself, give himself orders, obey, blame and punish himself; he can ask himself a question and answer it. We could even imagine human beings who spoke only in monologue; who accompanied their activities by talking to themselves. […] (PI, §243)

The point here is that naming or describing your own sensations from the position of an epistemic subject who does not have direct access to anything but those sensations cannot be a practice. In other words, a subject who only considers his or her private sensations cannot invent a practice of speaking about them, if such a practice must include the use of names to refer to sensations.

Why is this so? To keep things simple, we focus on Wittgenstein's argument for the impossibility of naming a sensation. Let us call the subject who does not have direct access to anything but her own sensations Joan. Now, it is important to note that Joan cannot talk about what causes her sensations or about the effect of her sensations on her behaviour until she can talk directly about her sensations. Then we have to imagine the way in which Joan could use a name to refer to a sensation. We take it that the sensation in this case is not a particular unrepeatable sensation but a reoccurring one. Two steps are involved. In the first step, Joan introduces a name for her sensation. In the second, Joan uses the name to refer to the same sensation. In Wittgenstein's own words:

> Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation. […] (PI, §258)

Joan cannot introduce the name "S" by using a description of the sensation for which the name stands because, in order to do this, she should talk about the qualities S has in common with other sensations to which she must be able to refer directly. Because we have to face such a case anyway, we should consider that S's case is similar to that. So the only alternative left is that "S" is introduced by an ostensive definition.

Wittgenstein's critique of ostensive definitions is well known, but that critique amounts only to the idea that a language cannot be learned starting with ostensive definitions. The main reason for that idea was that recognizing something as an ostensive definition requires that we already know how to take part in a series of linguistic practices. Here, however, we do not make use of that idea. To be clear about this, let us say that either Joan is in principle able to learn new words by ostensive definitions or she could invent herself the practice of introducing names by ostensive definitions. Were she to give a name to an object from her environment, she could take hold of it and write the name on it, for instance. The problem is that at this point Joan does not consider her environment but only her sensations. As a consequence, she must use a private ostensive definition. Instead of taking hold of an object or pointing at it, she tries to concentrate her attention on the sensation she wants to call "S". There is, however, an important difference between the two kinds of acts. Although pointing to something is a public act, concentrating your attention on a sensation is not. Public acts, in Wittgenstein's view at least, can succeed or fail and we have established criteria in order to recognize whether an act

does in fact succeed or not. If I try to point to something which my interlocutor cannot see, she can tell me so. Even if I am alone, I can think that I am pointing to (or waving at) another person and it is in principle possible for me to discover at a later time that I have failed to do so. A trick of the light made me think there was a person there when in fact there was not, etc.

I am taking one step further and considering the extreme case in which, after naming a small rock, one throws it away to a place full of similar small rocks, thus making it impossible to recognize it again. It could be convincingly argued even for such a case that the person in this case did succeed in introducing the name of that small rock by ostensive definition. The criteria of success for the definition and implicitly for the act of pointing at that small rock (or taking hold of it) are independent of the subsequent uses of the name introduced by the definition.

The case of a private ostensive definition is, however, different. The only guarantee Joan can have that she succeeded in pointing inwardly to her sensation is that her private act produces the expected result—the correct identification of the sensation in case in future cases. This makes the correctness of the private ostensive definition dependent on the future uses of the name thus introduced. However, if such a name is to be used to refer to a private mental object, the correctness of such a use can be established only by comparing it with the use of the same name when it was introduced. It is this circularity which makes it impossible to distinguish between correct and incorrect uses of a private name. As Wittgenstein puts it:

> Let us imagine a table (something similar to a dictionary) that exists only in our imagination. A dictionary can be used to justify the translation of a word X by a word Y. But are we also to call it a justification if such a table is to be looked up only in the imagination? - «Well, yes; then it is a subjective justification.» - But justification consists in appealing to something independent. - «But surely I can appeal from one memory to another. For example, I don't know if I have remembered the time of departure of a train right and to check it I call to mind how a page of the time-table looked. Isn't it the same here?» - No; for this process has got to produce a memory which is actually *correct*. If the mental image of the time-table could not itself be *tested* for correctness, how could it confirm the correctness of the first memory? (As if someone were to buy several copies of the morning paper to assure himself that what it said was true.) […] (PI, §265).

Another reason why Joan cannot name her sensations is that in order to do so she must have the concept of a sensation. For instance, we usually say that a feeling of uneasiness is not a sensation because it is not the direct result of our sense organs being affected by our environment. However, Joan cannot speak about her sense organs. In order to distinguish a feeling of uneasiness from a sensation she would have to talk only about qualia-like inner experiences (an experience such as "it feels as a feeling and not as a sensation"). In addition, any such qualia-like inner experiences would have to be distinguished from each other. This leads to an infinite regress. This is why Wittgenstein says:

> What reason have we for calling "S" the sign for a *sensation*? For "sensation" is a word of our common language, not of one intelligible to me alone. So the use of this word stands in need of a justification which everybody understands. […] (PI, §261).

The same objection could be raised when we imagine that, in order to introduce 'S' by a private ostensive definition, Joan performs an act which she calls 'directing my attention towards that sensation'. Calling a private act the act of directing one's attention towards something (in contrast, for instance, with cases in which "something gets one's attention" or "one gets interested in something" and so on) requires, in Joan's case, an inner qualia-like experience specific to the act of directing one's attention towards something. This experience must be such as to justify the application of the concept of "attention" to the act performed.

On my understanding, the same kind of objection already underlies Wittgenstein's reply to the attempt to justify the correctness of an application of 'S' by invoking a different sort of inner experience available by introspection—the belief that something is right:

> «Well, I *believe* that this is the sensation S again.» - Perhaps you *believe* that you believe it! […] (PI, §260)

What is Wittgenstein saying here? It is not that a person can hold a belief and yet be wrong, or at least mistaken with respect to the content of that belief being meaningful. Even if there were beliefs which could justify the application of a sign by themselves, we would still have the problem of recognizing something as such a belief. The correct application of 'S' would still depend, in this case, on the correct application of the concept of 'belief' to some private episode.

It could also be noted that, in order to *name* a sensation, Joan should have the concept of a name. The case which seems unproblematic with respect to establishing a connection between a name and a particular object is that in which I write the name on the object. However, writing a mark on an object could be part of a great variety of different practices. I could put a mark on an object to enable someone (perhaps myself) to recognise the object, or to enable someone to remember the object more easily (as a mnemonic device); I could use the name as a tag on a container, to indicate something about its content or as some visual aid when counting a range of objects ("I have counted up to this one"). More generally, it is in principle possible to use a mark as a visual aid for a lot of different activities—I mark the place where I want to make a hole in an object, I mark your height on the door frame, etc. If what distinguishes all these practices from that of introducing a mark in order to use it as a name is my intention to refer to the object at a later time by using the mark, we are facing the same problem: we have to apply the concept of *that particular intention* to some inner experience which we had when writing down the mark. In Wittgenstein's words:

> It might be said: if you have given yourself a private definition of a word, then you must inwardly *undertake* to use the word in such-and-such a way. And how do you undertake that? Is it to be assumed that you invent the technique of using the word; or that you found it ready-made?
>
> «But I can (inwardly) undertake to call THIS 'pain' in the future.» - But is it certain that you have undertaken it? Are you sure that it was enough for this purpose to concentrate your attention on your feeling? - A queer question. -" (PI, §262–3)

This recurrent theme points out to one moral: if we conceive mental episodes as private, then we cannot conceptualize them. A side point of the RFC was that we can have a concept only insofar as we have a practice, but we needed the PLA to understand why one could not invent a practice starting from the position of a subject who only has access to his or her own sensations. The foundation of knowledge, then, cannot consist of experiences conceived as sense data. This is, of course, not to say that one cannot conceive human experience in such a way that it would not consist of private mental episodes and it would have conceptual content.[2] Wittgenstein himself seems to offer us some suggestions in *On Certainty*.

Our problem is to understand what we do when we justify our everyday knowledge based on our experiences without talking of any "raw sensations". To see what Wittgenstein's suggestion could be, let us look at a simple case. Suppose we talk about the following belief:

(Chair) There is a chair in this room.

Now, the traditional empiricist's justification for (Chair) would perhaps be that she has being-in-this-room-like sensations and chair-like sensations. Wittgenstein's reaction to this is unchanged:

> An inner experience cannot shew me that I *know* something. Hence, if in spite of that I say, « I know that my name is…», and yet it is obviously not an empirical proposition,— (OC, §569)

No amount of psychological introspection can justify the belief that my name is X or the belief that (Chair). In this sense, (Chair) is not empirical.

According to Wittgenstein, the question of whether there is a chair in this room or not does not rise in any usual circumstances:

> […] I believe that there is a chair over there. Can't I be wrong? But, can I believe that I am wrong? Or can I so much as bring it under consideration? […] (OC, §173)

A situation could of course be imagined in which we would want to test for the presence of a chair in this room. In the same way:

> A mad-doctor (perhaps) might ask me «Do you know what that is?» and I might reply «I know that it's a chair; I recognize it, it's always been in my room». He says this, possibly, to test not my eyes but my ability to recognize things, to know their names and their functions. What is in question here is a kind of knowing one's way about. […] (OC, §355)

Nevertheless, in regular cases we do not *know* that there is a chair in this room. That is, we do not play the game of knowledge with respect to the chair in our room, but we *are certain* that there is a chair in the room, because this is an assumption of our actions involving the chair, including our verbal actions:

[2]McDowell (1994) is perhaps the most prominent attempt to do so. The connection of that project with PLA can be seen in McDowell (1989).

> Every language-game is based on words 'and objects' being recognized again. We learn with the same inexorability that this is a chair as that 2 × 2 = 4. (OC, §455)
>
> Children do not learn that books exist, that armchairs exist, etc. etc., - they learn to fetch books, sit in armchairs, etc. etc. […] (OC, §476)

In other words, (Chair) is a necessary condition for us performing some actions *as actions involving the chair in the room.*[3] If I sit in the chair in my room, for instance, (Chair) is assumed by my action, even if I do not express the belief (Chair). The following case is similar:

> Imagine a language-game «When I call you, come in through the door». In any ordinary case, a doubt whether there really is a door there will be impossible. (OC, §391)

The belief that "there really is a door" is assumed by the language-game imagined. That is, both the call to come in through the door and the action of coming in through the door are possible only if it is certain that there is a door. In another formulation:

> If I say «we assume that the earth has existed for many years past» (or something similar), then of course it sounds strange that we should assume such a thing. But in the entire system of our language-games it belongs to the foundations. The assumption, one might say, forms the basis of action, and therefore, naturally, of thought. (OC, §411)

Wittgenstein seems, however, to acknowledge the idea that believing (Chair) has nothing to do with being in a psychological state of belief towards the content of (Chair):

> […] Haven't I made the elementary mistake of confusing one's thoughts with one's knowledge? Of course I do not think to myself «The earth already existed for some time before my birth», but do I *know* it any the less? Don't I show that I know it by always drawing its consequences? (OC, §397)

Because (Chair) could be epistemically relevant (we could perhaps justify something within the game of knowledge by it) and we can be said to believe that (Chair) even if the belief in case does not consist in a particular mental state accessible by introspection, perhaps there is a sense in which it could be said that we *know* that (Chair):

> It is queer: if I say, without any special occasion, «I know» - for example, «I know that I am now sitting in a chair», this statement seems to me unjustified and presumptuous. But if I make the same statement when there is some need for it, then, although I am not a jot more certain of its truth, it seems to me to be perfectly justified and everyday. (OC, §553)

Now, if we accept that A is a necessary condition for B IFF B is a sufficient condition for A,[4] then from:

---

[3](Chair) is a necessary condition for us to perform intentional actions involving the chair in our room, i.e. actions under such descriptions as "Sitting on the chair in this room", "Moving the chair in this room (in such-and-such a way)", "Touching the chair in this room", "Testing to see whether the chair in this room supports a certain weight" and so on. See Anscombe (1963, §6).

[4]This could be disputed, of course. See, for instance, Sanford (1989, pp. 175–6).

(Nec) Our knowledge that (Chair) is a necessary condition for our actions involving the chair in this room (under such descriptions)

it follows that:

(Suff) Our actions involving the chair in this room (under such descriptions) are sufficient conditions for our knowledge that (Chair).

This could also give us some empiricist message: our experiences are the basis of our knowledge, but only if we conceive them as (constituted by) actions performed by us in our environment. It is my belief that this is the way in which we could understand Wittgenstein when he says:

> Giving grounds, however, justifying the evidence, comes to an end; - but the end is not certain propositions striking us immediately as true, i.e. it is not a kind of *seeing* on our part; it is our acting, which lies at the bottom of the language-game. (OC, §204)

To this, one could reply that the basis of our knowledge that (Chair) cannot consist of actions (involving the chair in the room) under some descriptions, because in order to conceive an action as "sitting in a chair" one has to know that "that is a chair", "one can sit on that", etc. According to this objection, even if we conceive our experience as being constituted from actions and not sense impressions, the same sort of circularity we were faced with in the case of private sensations still applies.

One escape route would be to note that sitting on chairs and talking about chairs are practices one can learn from others. These have to be shared practices. The communal character of language makes knowledge communal too:

> If experience is the ground of our certainty, then naturally it is past experience. And it isn't for example just my experience, but other people's, that I get knowledge from. […] (OC, §275)

This applies to scientific knowledge as well: the foundation of theoretical knowledge is our experience, formed by us acting in our environment. However, experience is not constituted by us performing "empirical actions" (touching, listening to, watching, observing, testing, experimenting, etc.) in isolation, but by performing such actions within shared practices.

What we get from Wittgenstein's view, by my understanding, is the suggestion that when we consider knowledge within the space of reasons,[5] the bottom line is represented by our empirical actions (considered as actions under descriptions). This is how we conceive experience as having conceptual content without talking about some pseudo-psychological entities which also have a conceptual content—perceptions.

In this sense, our beliefs about our environment share the same conceptual content with the actions performed by us in our environment. They represent knowledge because we use them to justify other beliefs, but in order to lose their

[5]See Sellars (1997, §36).

epistemic status they have to be challenged by the performance of other empirical actions, or perhaps by attempting to perform empirical actions which would justify such a belief and failing, as in the case where an attempt to sit on what seems to be a chair results in one falling down.

This is why in order to challenge a belief we have to accept other beliefs. For instance, I cannot challenge the belief that (Chair) if I do not believe that I have a body, because that belief is assumed by my "attempt to sit on something".

We are still left with a problem. How does a learned behaviour become an action under a description? The first thing we need to note is that this is not an epistemological problem. We need, of course, to make sure that the concepts we use to offer a natural explanation of this process do not lead us to any philosophical problems,[6] but an explanation must be possible in principle, because this is a fact of nature. We share some forms of behaviour and the capacity to learn some forms of behaviour with other animals and have evolved the ability to communicate (i.e. to perform the actions involved in speech) and to perform other intentional actions (i.e. actions under some descriptions) as well. The main trouble here is that we fear that at some point in the course of our natural explanation we have to switch from talking about natural events consisting of some kind of behaviour being displayed by human beings to talking about persons performing intentional actions within shared practices governed by rules. This is, however, far beyond the topic discussed here.

References

Anscombe, G. E. M. (1963). *Intention* (2nd ed.). Cambridge, MA, London: Harvard University Press.

Kripke, S. (1982). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.

McDowell, J. (1989). One strand in the private language argument. *Grazer Philosophische Studien, 33*, 285–303.

McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press).

Sanford, D. H. (1989). *If P, then Q: Conditionals and the foundations of reasoning*. London: Routledge.

Sellars, W. (1969). Language as thought and as communication. *Philosophy and Phenomenological Research, 29*(4), 506–527.

Sellars, W. (1997). *Empiricism and the philosophy of mind*. Cambridge, MA: Harvard University Press.

Wittgenstein, L. (1951). *Philosophical investigations*. Oxford: Blackwell.

Wittgenstein, L. (1969). *On certainty*. Oxford: Blackwell.

[6]Sellars's distinction between rules of action and rules of criticism is the sort of conceptual proposal which could be useful for a naturalist attempt to answer the question "How does a learned behaviour become an action under a description?" (see Sellars 1969: 508).

Author Biography

Gheorghe Ștefanov is Lecturer in the Faculty of Philosophy, University of Bucharest. He holds a Ph.D. in Philosophy from University of Bucharest in 2001, with a thesis on Wittgenstein's philosophy of language. Visiting Academic at Humboldt Universität, Berlin in 2002 (with a research grant offered by the Alexander von Humboldt Foundation). Has published articles and edited anthologies on Wittgenstein's philosophy, philosophy of logic and language, epistemology, philosophy of action and philosophy of technology. Author of "Fourteen Ideas of Ludwig Wittgenstein" (Humanitas Publishing House, 2013, in Romanian).

Part IV
Carnap

Carnap, Logicism, and Ontological Commitment

Otávio Bueno

Introduction

In a number of works, Rudolf Carnap provided ingenious and systematic approaches to the understanding of scientific knowledge (see, for instance, Carnap 1928, 1934, 1947, 1950a, b). As is well known, his proposals faced three important changes: (i) the early phenomenalism of the *Logische Aufbau der Welt* (Carnap 1928) was replaced by the physicalism of *Logische Syntax der Sprache* (Carnap 1934); (ii) the primarily syntactic account of *Logische Syntax* was then replaced by the semantic and modal views of *Meaning and Necessity* (Carnap 1947); and (iii) the deductivism that characterized Carnap's early work was eventually replaced by the probabilism of *Logical Foundations of Probability* (Carnap 1950a).[1] These shifts in philosophical perspective occurred against the background of the explicit use of mathematical and logical techniques, and they were articulated to preserve an essentially empiricist attitude towards science. The commitment to empiricism remained unchanged throughout Carnap's career.

I wish to thank Steven French and Jon Hodge for several discussions about Carnap's philosophy. Thanks are also due to Guillermo Rosado Haddock and Matthias Schirn for extensive and insightful comments on an earlier version of this work. Their comments led to substantive improvements.

[1]For a brief discussion of these shifts, see Jeffrey (1991). For a different non-phenomenalist interpretation of the early Carnap, see Haddock (2008, 2012).

O. Bueno (✉)
University of Miami, Coral Gables, USA
e-mail: otaviobueno@me.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_14

(According to Richardson (1998), before the *Aufbau*, and to some extent, even in the *Aufbau* itself, Carnap was not an empiricist, but a "critical conventionalist", who stressed the tie between the synthetic a priori and the logic of objective knowledge. On Richardson's view, a number of the themes and approaches developed by Carnap between 1921 and 1928 had neo-Kantian origins. In other words, the influence of neo-Kantians was crucial between Carnap's defense of his dissertation, *Der Raum* (Carnap 1922), and the publication of the *Aufbau* (Carnap 1928). From the *Aufbau* on, the commitment to empiricism remained constant.)

But in what sense is Carnap an empiricist? A crucial component of his empiricism comes down to the claim that we need to test our hypotheis about the world. Speaking about laws of nature (or, in Carnap's terminology, P-fundamental sentences), Carnap points out that:

> A sentence of physics, whether it is a P-fundamental sentence or an otherwise valid sentence or an indeterminate assumption (i.e., a premise whose consequences are investigated), is *tested*, in that consequences are deduced from it on the basis of the transformation rules of the language until one finally arrives at propositions of the form of protocol-sentences. These are compared with the protocol-sentences actually accepted and either confirmed or disconfirmed by them. If a sentence that is an L-consequence of certain P-fundamental sentences contradicts a proposition accepted as a protocol-sentence, then some alteration must be undertaken in the system. (Carnap 1934, p. 317)[2]

Empiricism comes then as the requirement of testability of scientific hypotheses. But of course there is more to empiricism than that. As Carnap points out:

> Empiricists are in general rather suspicious with respect to any kind of abstract entities like properties, classes, relations, numbers, propositions, etc. They usually feel much more in sympathy with nominalists than with realists (in the medieval sense). As far as possible they try to avoid any reference to abstract entities and restrict themselves to what is sometimes called a nominalistic language, i.e., one not containing such references. (Carnap 1950b, p. 205)

So, in Carnap's view, it is a crucial feature of empiricism that one should avoid ontological commitment to abstract entities. And given Carnap's use of mathematics throughout his program, he was aware of the difficulty that this use posed to empiricism.

But was Carnap really an empiricist? It might be argued that he could not possibly be an empiricist, for a crucial feature of Carnap's approach is a systematic attempt at providing a *neutral* stance between rival philosophical proposals, such as physicalism versus phenomenalism about the world, and platonism versus nominalism about universals (see Friedman 1999, pp. 206–210). And, of course, empiricism is *not* a neutral stance. In the *Aufbau*, Carnap is very clearly articulating a constructional system that could be adopted independently of the particular interpretation of science one favors. The resulting system is meant to be neutral between rival accounts of science, and so even those who have quite distinct ways of approaching scientific knowledge, such as realists and instrumentalists, can adopt

[2]For an illuminating discussion of this passage, see Friedman (1999, pp. 215–220).

it. The project is to provide the *form*, the overall *structure*, of scientific concepts, instead of favoring a particular philosophical interpretation of them.

Carnap was certainly searching for a neutral stance in philosophical disputes. But what motivated this stance was precisely the commitment to an empiricist view: an attempt to avoid ontological commitment to abstract entities (in particular, mathematical ones), and the requirement that scientific theories be testable. Note that the attempt is to *avoid the commitment* to the existence of abstract objects, rather than to *deny* that there are such objects. A *negative dogmatic* view about the existence of abstract entities would certainly fail to provide a neutral stance, since it would have to establish that there aren't such entities. However, an *agnostic* view, according to which we should "avoid any reference to abstract entities", provides exactly a neutral proposal. By avoiding reference to abstract entities, empiricists are neither committed to denying the existence of abstract objects, nor to claiming that there are such objects. They are simply neutral about the issue.

In the last two decades, increasing attention has been given to reexamine Carnap's philosophy. Richardson (1998), for example, has carefully studied the neo-Kantian origins of a number of Carnap's proposals in the *Aufbau*, and how Carnap adapted and reacted to important features of the neo-Kantian tradition. Friedman (1999) has also provided an insightful examination of the context in which Carnap's proposals emerged, examining, in particular, the nature of Carnap's empiricist view.[3] A rival interpretation of Carnap, resisting the neo-Kantian reading and highlighting the connections with Husserl, is advanced in Haddock (2008, 2012).

However, although Carnap's approach to the philosophy of mathematics plays some role in both Richardson's and Friedman's analyses, there is a conceptual shift in Carnap's view that has not received due attention: the strategies Carnap devised to avoid ontological commitment to mathematical entities. Given Carnap's empiricism, the question arises as to the nature of mathematical entities that are used throughout the development of his program—and since mathematical entities are abstract, the empiricist cannot simply believe in them. So how should one develop an empiricist approach to science, which heavily uses mathematical and logical resources, without having to believe in the existence of mathematical objects?

In this paper, I discuss and critically evaluate three crucial moves made by Carnap to accommodate mathematical talk within his empiricist program: (i) the "weak logicism" in the *Aufbau*; (ii) the combination of formalism and logicism in the *Logische Syntax*; and (iii) the distinction between internal and external question characteristic of Carnap's involvement with modality. As a result, the clear interplay between Carnap's philosophy of science and his work in the philosophy of mathematics will emerge, as well as some challenges that need to be overcome along the way.

[3]Particularly informative collections of essays on Carnap and other logical empiricists are Giere and Richardson (1996), Awodey and Klein (2004), and Friedman and Creath (2007).

Logicism in the *Aufbau*?

Before answering the question as to whether Carnap was a logicist, we should be clear about what is involved in a logicist position. Logicism comes in three dimensions (see Sainsbury 1979, pp. 272–274):

(a) Every mathematical truth can be expressed in a language whose expressions are logical (that is, whose expressions are formulated in terms of logic alone). In other words, all mathematical truths can be formulated as true logical propositions.

Note that a true logical proposition may not be a logical truth. For example, "There are at least three objects" is clearly a true logical proposition: it is true, and it can be formulated in terms of logic alone. However, it is not a logical truth. A logical truth, as Russell stressed, is "true in virtue of its form" (Russell 1903, p. xvi), and one typically requires that it be necessarily true.

(b) Every true logical proposition, which is a translation of a mathematical truth, is a *logical truth*. (Given that a true logical proposition may not be a logical truth in general, it becomes clear that (b) is stronger than (a).)
(c) Every mathematical truth, after being formulated as a logical proposition, can be *deduced* from a small number of logical axioms and rules.

As Sainsbury indicates, logicism is often presented as the conjunction of (a) and (c). For example, in *The Principles of Mathematics*, Russell claims:

> All pure mathematics deals exclusively with concepts definable in terms of a very small number of fundamental logical concepts, and […] all its propositions are deducible from a very small number of fundamental logical principles. (Russell 1903, p. xx)

It is important to distinguish (b) and (c), at least with hindsight, given Gödel's incompleteness theorem (Sainsbury 1979, p. 273). Roughly speaking, no consistent logical system, which is strong enough to formulate arithmetic, allows the derivation of all mathematical truths. So the logicist who only asserted (a) and (c) would have to revise (c), and allows that only a *substantial amount* of mathematical truths are derivable. (Let us call (c′) this revised version of (c).)[4] However,

[4]Of course, (c′) is weaker than (c), but despite being somewhat vague, it is strong enough to be taken as a logicist proposal. After all, we are still able to claim that the mathematical propositions derived from the logicist's stock are *logical* propositions. With hindsight, Gödel's incompleteness result shouldn't be surprising for the logicist. After all, due to Gödel's result, the logicist's crucial tool for the reformulation of mathematics—higher-order logic—is ultimately incomplete. As is well known, Frege used *second-order* logic in his logicist program whereas Russell employed *type theory*, both of which are incomplete in the standard semantics. (Of course, they become complete if we introduce Henkin models; see Shapiro (1991). But in this case it is unclear that we are really dealing with second-order logic in its proper form; see Bueno (2010).) Perhaps Frege's logicism, when restricted to arithmetic, can be interpreted in terms of the conjunction of (a) and (b). Given the role played by Hume's principle, and what follows from it, (c) may be needed as well. However, this is not the place to address the delicate issues raised by trying to make sense of Frege.

with (c′), the logicist is in no position to distinguish mathematical truths from true logical propositions that are *not* mathematical truths. In order to indicate that the mark of mathematical truth is *logical* truth, the logicist has to assert (b). And that is why he or she has to keep the three dimensions of logicism distinct. Ultimately, the logicist asserts (a), (b) and (c′).

Now there is an additional aspect of logicism that is worth mentioning here. With the conjunction of (a), (b) and (c′), the logicist can claim that mathematics can be reformulated as a deductive-definitional development of logic (see Bohnert 1975, p. 184). But this leaves open the issue of whether logic, and hence mathematics, is *analytic*.[5] If we add to (a)–(c′) the claim (d) that logic is indeed analytic, we have what Bohnert (1975) calls *strong logicism*.

Now, Carnap was certainly *not* a logicist in the strong sense (Bohnert 1975). He clearly separated geometry, physically interpreted, from mathematics. Thus, in order to interpret geometry physically, we need to go beyond purely logical considerations. This requires a revision of (a), since certain geometrical truths aren't true logical propositions (given that they introduce physical notions). Carnap also adopted an extensionalist approach (at least in the *Aufbau* and in *Logische Syntax*) in contrast with the intensional stance favored by Frege and Russell. In particular, Carnap rejected Russell's reducibility axiom and adopted the simple theory of types instead of the ramified theory (Carnap 1931). In this respect, the conceptual framework Carnap used was not the one developed by the first logicists. Finally, following Russell, Carnap granted that the axioms of infinity and choice were *not* logical truths (Carnap 1931). He took these principles, just as Russell did, as hypotheses (see Russell 1919; Sainsbury 1979, pp. 305–307). As a result, (b) can't be asserted in general, since there are translations of mathematical truths into a logical language (such as the axiom of infinity) that are *not* logical truths. Thus, according to Bohnert, until the early 1930s, Carnap's logicism

> was of the more modest, conditional kind which asserted that while the concepts of mathematics were reducible by definition to logical ones, the truths of mathematics were deductively reducible only to logic plus axioms of infinity and choice. (Bohnert 1975, p. 184)

In fact, in his 1931 discussion of logicism, Carnap presents logicism as "the thesis that mathematics is reducible to logic, hence nothing but a part of logic" (Carnap 1931, p. 41). He then splits the logicist thesis into two parts: (a) "the

[5]It should be noted that, at least until 1903, Russell thought that logic was *synthetic*. On his view, given that Kant established that arithmetic is synthetic, and given Frege's logicist program of reducing arithmetic to logic, we should conclude that logic is also synthetic, rather than that arithmetic is analytic. As he points out in *The Principles of Mathematics*: "Kant never doubted for a moment that the propositions of logic are analytic, whereas he rightly perceived that those of mathematics are synthetic. It has since appeared that logic is just as synthetic as all other kinds of truth" (Russell 1903, p. 457, see also 1912, Chap. VIII; for a discussion, see Dreben 1990, pp. 86–87, and 93, note 67).

concepts of mathematics can be derived from logical concepts through explicit definitions", and (b) "the *theorems* of mathematics can be derived from logical axioms through purely logical deduction" (Carnap 1931, p. 41). (I will return to this paper below.)

Despite these considerations, it is usual to read Carnap's *Aufbau* as an extension of logicism into the philosophy of science. After all, Carnap arguably uses logicist techniques to provide an interpretation of science. To some extent, this is the way in which Carnap himself presented his approach. In the *Aufbau*, Carnap discusses the method to be used in the constructional system, which was meant to provide "the analysis of reality with the aid of the theory of relations" (Carnap 1928, p. 7). The main point is to show how all the necessary concepts to formulate science can be articulated in a constructional system, just as logicists have articulated arithmetical concepts. Not surprisingly, the example used by Carnap to illustrate a constructional system is provided by arithmetic:

> A constructional system of arithmetical concepts can be established by deriving or "constructing" step-by-step (through chains of definitions) all arithmetical concepts from the fundamental concepts of number and immediate successor. (Carnap 1928, pp. 6–7)

This is the kind of construction that Carnap is trying to extend to the field of science. As Carnap indicates, the method to be used is basically the theory of relations put forward by Russell, "based on the pioneer work of Frege" (Carnap 1928, pp. 7–8).[6]

The association of Carnap's approach with logicism also comes from the famous symposium on the foundations of mathematics in Königsberg in 1930 where Carnap discussed the logicist approach (Carnap 1931), whereas Heyting supported the intuitionist view, and von Neumann provided a defense of formalism.[7]

However, these two references to logicism, in the *Aufbau* and in the 1931 paper, shouldn't lead us to believe that Carnap wholeheartedly adopted a logicist proposal. We have already noticed a number of reservations that Carnap had to the standard formulations of logicism. And in the context of the 1931 paper, after indicating some advantages of logicism (in the formulation of mathematical concepts through explicit definitions), Carnap didn't hesitate to indicate its limitations. They arise in the context of the derivation of mathematical theorems. In Russell's type-theoretic approach, the reducibility axiom has to be introduced alongside axioms of infinity and choice.[8] Now, the latter two axioms are existential in character, and hence

[6]It is an issue in the interpretation of Carnap's work, whether in the *Aufbau* Carnap was developing a *constructional* or a *constitutive* system, depending, respectively, on the neo-Kantian or Husserlian reading of the work that is advanced. For the purposes of the present paper, I won't take sides on this issue, since the points I want to highlight will go through on either reading. I've used the term 'constructional' for convenience only.

[7]English translations of these papers can be found in Benacerraf and Putnam (1983, pp. 41–65).

[8]Here are Carnap's own formulations of these axioms: "The axiom of infinity states that for every natural number there is a greater one. The axiom of choice states that for every set of disjoint non-empty sets, there is […] a set that has exactly one member in common with each of the member sets" (Carnap 1931, p. 44).

cannot be presented as "logical axioms", since "logic deals only with possible entities and cannot make assertions about whether something does or does not exist" (Carnap 1931, p. 45).[9]

Moreover, logicism as such, at least as formulated by Frege, is not ontologically innocuous, and it doesn't provide a straightforward "nominalization strategy" for mathematics, in the sense of making abstract mathematical notions acceptable to an empiricist. After all, certain concepts are required for the logicist reconstruction of arithmetic, and concepts in general, as understood by the logicist, are abstract. Therefore, it is not by chance that Carnap assumes a blend of logicism *and* conventionalism in the *Aufbau* (see below): in this way, he thinks he is able to avoid ontological commitment to mathematical entities (for further discussion of nominalist versions of logicism, see Bueno 2001).

Not surprisingly, Carnap's logicism is not "pure", in the sense that he doesn't countenance Frege's distinction between concept and object (Carnap 1928, p. 10), which is crucial for Frege's construction of arithmetic in terms of logic.[10] In Carnap's own words:

> Since we always use the word "object" in its widest sense […], it follows that to every concept there belongs one and only one object: "its object" (not to be confused with the objects that fall *under* the concept). In opposition to the customary theory of concepts, it seems to us that the generality of a concept is relative, so that the borderline between general and individual concepts can be shifted, depending on the point of view […]. Thus, we will say that even general concepts have their "objects". (Carnap 1928, p. 10)

Of course, in the passage above, the remark about objects *falling under a concept* clearly refers to Frege's account, which is not the one adopted by Carnap. It is not by chance that Carnap doesn't advocate this distinction, since it brings a

[9]At this point, Russell's move is well known. On his view, certain mathematical statements only have a conditional form. If a proof of a mathematical statement M requires the axiom of infinity (AI) or of choice (AC), we should take M as a conditional of the form: AI $\rightarrow M$ or AC $\rightarrow M$, respectively. And this conditional can be derived from the axioms of logic (see Carnap 1931, p. 45). As it stands, the problem with this move is that it is inadequate as a general strategy for mathematics. Unless we establish the *consistency* of the antecedent, the consequent can be obtained trivially, in which case both M and its negation are derivable from the axioms of logic.

[10]The concept/object distinction is, of course, crucial for Frege's ontology. In drawing it, Frege relies on certain syntactic criteria. As Matthias Schirn noted (in personal communication), the distinction is drawn, in the first place, independently of Frege's foundational or logicist concerns, although it proved to be quite essential for Frege's logicist project. Frege conceived numbers as objects, identifying all numbers with logical objects of a fundamental and irreducible kind: value-ranges of functions (including extensions of concepts and extensions of relations). He added his theory of value-ranges to second-order logic due to the need for introducing objects purely logically. Frege's conception of numbers was also decisive for his proof of the infinity of the series of cardinal numbers. (For a thorough and illuminating examination of Frege's foundational work with special attention to Hume's Principle, see Schirn 2015).

serious ontological problem for the empiricist: as noted, concepts—in Frege's sense—are abstract entities and cannot be accepted in the empiricist's ontology.[11]

The alternative proposed by Carnap is to provide a *conventionalist* reading of the mathematics used in his program,[12] while sticking to a *weak logicist* reconstruction of it (Carnap 1928, pp. 177–178). On his view, mathematical terms have no reference, since they are logical constructions, obtained by explicit definition. In his own words:

> It is important to notice that the logical and mathematical objects are not actually objects in the sense of real objects (objects of the empirical sciences). *Logic* (*including mathematics*) *consists solely of conventions* concerning the use of symbols, *and of tautologies* on the basis of these conventions. Thus, the symbols of logic (and mathematics) do not designate objects, but merely serve as symbolic fixations of these conventions. Objects in the sense of real objects […] are only the basic relation(s) and the objects constructed therefrom. (Carnap 1928, p. 178)

This passage clearly illustrates the two crucial components of the *Aufbau*'s nominalization strategy: a conventionalist understanding of mathematics, and the use of logicist constructions (the tautologies on the basis of the conventions) to resist the claim that mathematical terms refer. Carnap is objecting to the idea that logical and mathematical objects should be conceptualized as objects at all, given that, on his view, logic and mathematics are concerned only with conventions regarding the use of symbols. But consider what results from the application of the relevant conventions; do the resulting *things* (to use a neutral term) exist or not? It's unclear that a compelling answer emerges in this case. Perhaps Carnap would prefer to get rid of the question altogether—and we will see how he eventually will try to do that in the 1950s with the introduction of internal and external questions (Carnap 1950). But at this point, how successful is Carnap's strategy?

Clearly the conventionalist component is not enough to guarantee that mathematical and logical terms lack reference. Conventions can be established for and applied to both existing and non-existent things, and it seems only by *fiat* that in the case of mathematics they would do the nominalist trick. Even if it were true that logic and mathematics consisted only in conventions regarding the use of symbols, and in tautologies obtained from these conventions, it doesn't follow that what these symbols stand for are not abstract objects, particularly given the existential content of so many mathematical theorems, which state the existence of infinitely

[11]It should be noted that there are a number of puzzling features in Carnap's quotation above (the one from his 1928, p. 10). I'll mention two. First, it is unclear exactly why "to every concept there belongs one and only one object: 'its object'"—even if 'object' is used in a very wide sense, and thus includes Fregean functions, concepts and relations. Presumably Carnap would recognize first-level concepts under which more than one object falls, although that doesn't help to clarify the matter. Second, the distinction between general and individual concepts is not specified, and thus it is unclear exactly why the generality of a concept is supposed to be relative. I won't pursue these concerns further here, but would like to thank Matthias Schirn for pressing me on these issues.

[12]In this respect, Richardson's point about Carnap's conventionalism in the *Aufbau* is very well taken (see Richardson 1998).

many primes, simply-connected spaces, or solutions to a variety of differential equations.

It should come as no surprise then that Carnap himself felt the need for an additional nominalization strategy. In his 1931 discussion of logicism, published just three years after the *Aufbau*, he once again indicated the role of logical constructions in mathematics. However, he stressed that the deductions and definitions provided are "carried through formally as in a pure calculus, i.e., without reference to the meaning of the primitive symbols" (Carnap 1931, p. 52). And this remark added something substantially new to what was argued in the *Aufbau*. There is a clear move, on Carnap's side, towards the *formalist* camp. On his view, if all we have in mathematics is a "pure calculus", with no reference to the "meaning of the primitive symbols", we could be talking about any fiction whatsoever, and no existence assumption is then required. This is a leading idea in Carnap's work since the 1930s, and it is the driving force behind Carnap's next important publication: *Logische Syntax der Sprache* (Carnap 1934).

Logicism in the *Logische Syntax*?

The "mixture" of formalism and logicism is a characteristic feature of *Logische Syntax*. And it is not by chance that the book was taken by some critics as a "surrender of logicism to formalism" (Beth 1963, p. 476). As Beth acknowledges, though, the logicist component is still strong, since we find, in *Logische Syntax*, concepts that

> though defined in a purely formal fashion, are clearly inspired by a non-formal interpretation which, if made manifest, would imply a return to Frege's logicism. I think that it is even possible to show that, in the absence of such a non-formal, intuitive, interpretation, the whole edifice of *Logical Syntax* would miss its purpose. (Beth 1963, p. 477)

In order to support this claim, Beth (1963, pp. 477–479) argues that a variant of the Löwenheim-Skolem theorem can be generated in Carnap's framework, and he concludes that Carnap has to assume an intuitive, non-formal, interpretation as a crucial component of his system. In other words, Carnap needs an *intended* interpretation for his formal construction—but this is a return to logicism. After all, a logicist formulation of arithmetic provides exactly this: a particular interpretation of arithmetic that preserves the intended meaning of arithmetical terms.

Beth's argument goes as follows:

> Following the lines of Gödel's argument, Carnap constructs a certain sentence W_{II} which belongs to his Language II [the language of classical mathematics] and which constitutes the arithmetisation of a certain sentence in the syntax of Language II expressing the consistency of Language II. It can be shown that, if Language II happens to be consistent—which will be assumed throughout—then W_{II} cannot be proven in Language II, though, according to the intended intuitive interpretation of Language II, W_{II} must, of course, be true.

> Hence if non-W_{II}, the negation of W_{II}, is introduced as a new axiom, so as to extend Language II into a stronger system II*, then II* is still consistent. On account of Henkin's completeness theorem for the theory of types, which can be proven within II, II admits a certain model M*, which, of course, must be different from the intuitive "standard" model M for II. (Beth 1963, pp. 477–478)

Before proceeding, Beth introduces a hypothetical logician, Carnap*, whose intuitions are in accordance with model M*. Beth then examines Carnap*'s view of the situation:

> It seems reasonable to suppose that, for Carnap*, all theorems of II are intuitively true. Now these theorems include the arithmetisations of the theorems of a certain extension S* of the Syntax of II; arithmetisation is an interpretation, hence a translation; it follows that, for Carnap*, the theorems of S* must be intuitively true. As II* contains *non*-W_{II}, it follows that for Carnap* Language II is inconsistent. On the other hand, S* contains Henkin's theorem, hence, for Carnap*, Language II, and a fortiori II*, cannot admit a model. However, Carnap* was supposed to be endowed with the intuitive model M*! (Beth 1963, p. 478)

Given that Carnap and Carnap* would always agree on conclusions that depend only on formal considerations, this argument seems to indicate an incoherence in Carnap's proposal. But is this really so?

According to Beth, there is a way out, but this means giving up a purely formalist view, and introducing a crucial role for logicism:

> The solution of the paradox is as follows. We made an error in supposing that, for Carnap*, all theorems of II* are intuitively true; for Carnap*'s set of all theorems of II* does not coincide with *our* set of all theorems of II*. Roughly speaking, Carnap*'s set contains *more* theorems than ours, and for some of these additional theorems M* is not a model. (Beth 1963, p. 478)

Beth's conclusion is then clear:

> The above considerations, which are only variants of the Skolem-Löwenheim paradox, suggest strongly that, if arguments as contained in *Logical Syntax* serve a certain purpose, this can only be the case on account of the fact that they are interpreted by reference to a certain presupposed intuitive model M. Carnap avoids appeal to such an intuitive model in the discussion of Language II itself, but he could not avoid it in the discussion of its syntax; for the conclusions belonging to this syntax would not be acceptable to Carnap*, though Carnap and Carnap* would, of course, always agree with respect to those conclusions which depend exclusively on formal considerations. (Beth 1963, pp. 478–479)

So what we have in *Logische Syntax* is ultimately a sophisticated combination of formalism and logicism. This is a feature that Carnap acknowledges as one of the *advantages* of his proposal:

> The formalist view is right in holding that the construction of the [mathematical] system can be effected purely formally, that is to say, without reference to the meaning of the symbols; that it is sufficient to lay down rules of transformation, from which the validity of certain sentences and the consequence relations between certain sentences follow; and that it is not necessary either to ask or to answer any questions of a material nature which go beyond the formal structure. (Carnap 1934, p. 326)

However, Carnap proceeds, the formalist construction of a calculus is not enough. After all,

> this calculus does not contain all the sentences which contain mathematical symbols and which are relevant for science, namely those sentences which are concerned with the *application of mathematics*, i.e. synthetic descriptive sentences with mathematical symbols. (Carnap 1934, p. 326)

At this point, the *logicist* enters. According to Carnap (1934, p. 326), the sentence "Charles and Peter are in this room now and no one else" doesn't entail the sentence "There are now two people in this room", given the usual way in which formalists construct their calculus, given that no meaning considerations are presupposed in the inference rules, and no reference to mathematical objects is available. However, the derivation does go through with the use of the *logicist* system, given Frege's definition of "2". (Remember, however, that Frege's definition of numbers is articulated in terms of an *abstract* notion of concept.) In other words, it is the *combination* of formalism and logicism that provides an adequate framework:

> For, on the one hand, the procedure [of characterizing mathematics] is a purely formal one [in conformity with formalism], and on the other, the meaning of the mathematical symbols is established [as the logicist requires] and thereby the application of mathematics in actual science is made possible, namely, by the inclusion of the mathematical calculus in the total language. (Carnap 1934, pp. 326–327; italics omitted)

In this way, Carnap concludes:

> the task of the logical foundations of mathematics is not fulfilled by a metamathematics (that is, by a syntax of mathematics) alone, but only by a syntax of the total language, which contains both logico-mathematical and synthetic sentences. (Carnap 1934, p. 327; italics omitted)

However, is this combination of logicism and formalism satisfactory? I don't think so. Carnap is certainly right in stressing that even if the formalist successfully reformulates mathematics as an uninterpreted calculus, the question of its application has to be faced. And the formalist at least owes us an explanation of why a meaningless calculus can be so useful in the description of the physical world. At this point, Carnap will probably move back to the logicist approach, arguing that the usefulness of mathematics can be accommodated in terms of the way logicists introduce concepts (such as numbers), and indicating that the latter are crucial for the application of mathematics. The problem with this move comes at the ontological front, since the concepts introduced by the logicist are abstract, and thus they do not mesh with Carnap's empiricism. So, as it stands, logicism—even in the weak form provided by Carnap—is ontologically inflationary, and formalism, despite being *prima facie* ontologically acceptable, is unable to accommodate the application of mathematics, which is similarly crucial for empiricism.

We can see why the combination of these two strategies seemed so attractive to Carnap, but together they don't seem to generate a stable proposal. If mathematics is just a calculus, it is unclear how it can be so successful in applications; and if

mathematics is obtained via a logicist reconstruction, it is unclear how it can be made compatible with empiricism. Either way, difficulties emerge for Carnap's overall proposal. Is there a way out?

A Logicist Involvement with Modality?

In light of Gödel's theorems, many thought that formalism was no longer a viable alternative in the foundations of mathematics. Carnap also realized the need for semantic notions in the analysis of science: this was probably the most adequate way to introduce the notion of analyticity (Carnap 1947; see also Bohnert 1975). This led Carnap to articulate a different nominalization strategy of mathematics, in terms of his well-known demarcation between internal and external questions (Carnap 1950b). This strategy suggests a comprehensive way in terms of which the empiricist can use abstract notions—not only in semantics, but also in the philosophy of science—to provide an interpretation of theories.

In order for questions such as 'Are there numbers, properties, or universals?' to be properly formulated, they need to be raised within a particular framework: a formal theory that characterizes the concepts invoked in the questions themselves. The proposed answers then become internal to what follows from the characterizing features of the framework: they are analytically true or empirically true, depending on the framework under consideration and the question being asked. However, if the question is raised independently of any framework, then it is unclear what is being asked: after all, a framework is needed to specify the concepts invoked in the question. Such external (framework-independent) questions are therefore non-cognitive: they are concerned with the existence (or not) of certain objects *without properly specifying them*. External questions ultimately require a pragmatic choice among different frameworks, which provide distinct characterizations of the concepts in question.

This strategy, however, being so comprehensive, doesn't yield a particularly insightful account of mathematics. There is nothing *specific* to mathematics that the external/internal strategy highlights. It is too general, and if successful, would provide a nominalization strategy not only for mathematical notions, but also for semantic and theoretical concepts. But does it succeed? It's not clear that it does. Within a framework, such as one that characterizes numbers and other abstract objects, entities of this kind exist. But this supports a form of realism about these objects, albeit one that is framework dependent. This conclusion, however limited, is not one available to the empiricist, though, who will need to employ a different framework in which abstract objects don't exist. But how can the applications of mathematics be accommodated within such a framework? To answer this question, one would need to nominalize mathematics. But that's precisely what the internal/external distinction was meant to achieve in the first place—one is then back to square one. As a result, this strategy doesn't solve the problem of the status of mathematical objects for an empiricist.

Given the inconclusive nature of the external/internal dichotomy with regard to the status of mathematical objects, it is not surprising that, in 1958, Carnap still maintained the blend of formalism and logicism as a nominalization of mathematics. This becomes clear in his paper "Observation Language and Theoretical Language" (Carnap 1958). He starts the paper by briefly recalling the main ideas of type theory, and he notes that the whole of classical mathematics can be recaptured type-theoretically; in particular, one can reconstruct the various sorts of numbers (from natural to complex numbers), relations between such numbers, function of numbers etc. Carnap then claims:

> But *we make no ontological assumption about the existence* (*in a metaphysical sense*) *of such objects*. We only introduce expressions which follow definite syntactical rules. In other words, *we specify a calculus*. The syntactical rules, for example primitive sentences and rules of inference, are formulated in a syntactical metalanguage. […] For these rules, a metalanguage with only an elementary logic […] is sufficient. (Carnap 1958, p. 77; the italics are mine)

We have already seen Carnap using this formalist argument back in 1931. And his conclusion, after 27 years, is equally surprising:

> Such a structure of mathematics corresponds to the procedures recommended by Hilbert and carried out by him, together with Bernays. *These methods can also be accepted by those of an intuitionistic, constructivist, or even a nominalist orientation, since the elementary syntactical metalanguage satisfies the requirements which these orientations impose on a meaningful language*. (Carnap 1958, p. 77; the italics are mine)[13]

But the same problems that faced the blend of logicism and formalism before still apply here. As opposed to what is allowed by the external/internal dichotomy, Carnap makes here a *definite* ontological claim, *denying* the existence of mathematical objects, given his claim that nominalists can accept the methods in question. However, given the considerations presented at the end of Section "Logicism in the *Logische Syntax*", the resulting combination of formalism and logicism doesn't succeed: the formalist component doesn't accommodate the application of mathematics, and the logicist component is not compatible with empiricism. Given that the various nominalization strategies advanced by Carnap were articulated, in part, to maintain empiricism, some more additional work is needed.

To be fair to Carnap, he does introduce a new twist with this argument. He claims that we don't have to assume the existence of the mathematical objects formulated in type theory. For type theory is only a calculus, and the syntactic rules of this calculus are formulated in a metalanguage that only requires an elementary logic (which is, thus, substantially weaker than type theory itself). But why would this provide a nominalization strategy for mathematics? Even if, as Carnap claims, intuitionists, constructivists and nominalists could adopt the elementary metalanguage, it doesn't follow that mathematical objects don't exist. Even if mathematical claims were *not* made in the metalanguage, but only in the object language, if

[13]It is worth noting that, in this type-theoretic context, Carnap argues that analyticity can be defined in terms of Ramsey sentences (see Carnap 1958, pp. 81–84).

intuitionists, constructivists and nominalists use the latter, given its type-theoretic content and the amount of mathematics that can be formulated with it, they will be committed to abstract entities—unless a suitable nominalization of mathematics is advanced. There is no way out.

Conclusion

As I argued in this paper, the answer to the question 'Was Carnap a Logicist?' is complex. Throughout his career, Carnap was never a straightforward logicist. Surely there were logicist features in his thinking, but these were articulated in combination with other theoretical approaches. As noted above, Carnap developed three different combinations:

1. In the *Aufbau* and in the 1931 paper, Carnap didn't adopt the then standard formulations of logicism: he didn't countenance Frege's distinction between concept and object, nor did he accept Russell's ramified type-theory. Moreover, he combined the resulting weak version of logicism with a thoroughly *conventionalist* outlook.
2. Dissatisfied with the answer provided by the combination of conventionalism and weak logicism to the (nominalist) status of mathematics, Carnap shifted to a *formalist* approach in the *Logische Syntax*. The logicist component is also found here, in combination with the claim that mathematics is basically an uninterpreted calculus.
3. But Carnap soon realized the limitations of a purely syntactic account, and developed a comprehensive semantic approach. With the introduction of semantic notions, a corresponding strategy at the ontological front was articulated, with Carnap's distinction between internal and external questions. Although the distinction didn't provide a specific nominalization strategy for mathematics, it allowed the introduction of semantic notions in a thoroughly empiricist way. With regard to the issue of nominalism, however, Carnap would still maintain the *formalist* attitude put forward in *Logische Syntax*, taking mathematics as essentially an uninterpreted calculus (Carnap 1958). The logicism found at this level is still weak, helping to reconstruct and reformulate mathematics in terms of the simple theory of types.

As we saw, however, the combination of formalism and logicism is ultimately unsuccessful. After all, if mathematics is taken as a mere calculus, as the formalist wants, it is a mystery how it can be applied so successfully to science. And if mathematics is reconstructed according to logicist guidelines, it is unclear that it is compatible with an empiricist view. Despite Carnap's remarks to the contrary, it seems that the attempt at putting together weak logicism and formalism faces, in the end, a delicate predicament.

References

Awodey, S., & Klein, C. (Eds.). (2004). *Carnap brought home: The view from Jena*. Chicago and La Salle: Open Court.

Barrett, R., & Gibson, R. (Eds.). (1990). *Perspectives on Quine*. Oxford: Blackwell.

Benacerraf, P., & Putnam, H. (Eds.). (1983). *Philosophy of mathematics: Selected readings* (2nd ed.). Cambridge: Cambridge University Press.

Beth, E. (1963). Carnap's Views on the advantages of constructed systems over natural languages in the philosophy of science. In Schilpp (Ed.), pp. 469–502.

Bohnert, H.G. (1975) Carnap's logicism. In Hintikka (Ed.), pp. 183–216.

Bueno, O. (2001). Logicism revisited. *Principia, 5*, 99–124.

Bueno, O. (2010). A defense of second-order logic. *Axiomathes, 20*, 365–383.

Carnap, R. (1922). *Der Raum. Ein Beitrag zur Wissenschaftslehre*. *Kant-Studien* Ergänzungsheft 56. Berlin: Reuther und Reichard.

Carnap, R. (1928). *The logical structure of the world* (R.A. George, Trans. German). Berkeley: University of California Press, 1969.

Carnap, R. (1931) The logicist foundations of mathematics. *Erkenntnis 2*, 91–105. (English Trans. Putnam, E. & Massey, G. (1983) in Benacerraf and Putnam (Eds.), pp. 41–52).

Carnap, R. (1934) *The logical syntax of language* (A. Smeaton, Trans. German). London: Kegan Paul Trench, Trubner and Co., 1937.

Carnap, R. (1947). *Meaning and necessity* (2nd ed. 1956). Chicago: University of Chicago Press.

Carnap, R. (1950a). *Logical foundations of probability*. Chicago: University of Chicago Press.

Carnap, R. (1950b). Empiricism, semantics, and ontology. *Revue International de Philosophie 4*, 20–40. (Reprinted in Benacerraf and Putnam (eds.) [1983], pp. 241-257, and in the 1956 edition of Carnap [1947], pp. 205-221).

Carnap. R. (1958) Observation language and theoretical language. *Dialectica 12*, 236–248. (English Trans. H.G. Bohnert, published in Hintikka (Ed.) [1975], pp. 75-85).

Costreie, S. (Ed.). (2015). *Early analytic philosophy: New perspectives on the tradition*. Dordrecht: Springer.

Dreben, B. (1990). Quine. In Barrett and Gibson (Eds.), pp. 81–95.

Friedman, M. (1999). *Reconsidering logical positivism*. Cambridge: Cambridge University Press.

Friedman, M., & Creath, R. (Eds.). (2007). *Cambridge Companion to Carnap*. Cambridge: Cambridge University Press.

Giere, R., & Richardson, A. (Eds.). (1996). *Origins of logical empiricism*. Minneapolis: University of Minnesota Press.

Haddock, G. R. (2008). *The Young Carnap's Unknown Master: Husserl's Influence in Der Raum and Der logische Aufbau der Welt*. Hampshire: Ashgate.

Haddock, G. R. (2012). *Against the current*. Berlin: Ontos Verlag.

Hintikka, J. (Ed.). (1975). *Rudolf Carnap, logical empiricist*. Dordrecht: D. Reidel Publishing Company.

Jeffrey, R. (1991). After Carnap. *Erkenntnis, 35*, 255–262.

Richardson, A. W. (1998). *Carnap's construction of the world: The Aufbau and the emergence of logical empiricism*. Cambridge: Cambridge University Press.

Russell, B. (1903) *The principles of mathematics* (2nd ed. 1937). London: Routledge.

Russell, B. (1912). *The problems of philosophy*. London: Williams and Norgate.

Russell, B. (1919). *Introduction to mathematical philosophy*. London: Routledge.

Sainsbury, M. (1979). *Russell*. London: Routledge and Kegan Paul.

Schilpp, P. A. (Ed.). (1963). *The philosophy of Rudolf Carnap*. La Salle: Open Court.

Schirn, M. (2015) On the nature, status, and proof of Hume's principle in Frege's logicist project. In Costreie (Ed.).

Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*. Oxford: Clarendon Press.

Author Biography

Otávio Bueno is Professor of Philosophy and Chair of the Philosophy Department at the University of Miami. His research concentrates in philosophy of science, philosophy of mathematics, philosophy of logic, and epistemology. He has published widely in these areas in journals such as: *Noûs*, *Mind*, *British Journal for the Philosophy of Science*, *Philosophical Studies*, *Philosophy of Science*, *Synthese*, *Journal of Philosophical Logic*, *Studies in History and Philosophy of Science*, *Analysis*, *Studies in History of Philosophy of Modern Physics*, *Erkenntnis*, *Monist*, *Metaphilosophy*, *Studies in History and Philosophy of Biological and Biomedical Sciences*, *Ratio*, *History and Philosophy of Logic*, and *Logique et Analyse*. He is the author or editor of several books, and editor in chief of *Synthese*.

Frege the Carnapian and Carnap the Fregean

Gregory Lavers

Introduction

Many philosophers see Frege as the prototypical realist in the philosophy of mathematics. This realism is understood as a metaphysical position. Carnap, on the other hand, famously rejected metaphysics. Apart from their connection with logicism, their positions on the foundations of logic and mathematics may seem diametrically opposed. However, this caricatured picture of the relation between their philosophical positions is highly misleading.[1]

In the last twenty years or so, after a period of neglect, Carnap's philosophy has enjoyed renewed interest; so much so that he is now often placed, along with Frege, Russell and Wittgenstein, as a defining member of the early analytic tradition. Despite this situation, little has been written about the relationship between the views of Carnap and Frege. While biographical details of their interaction are quite well known, the relation between their philosophical positions has received less attention than it deserves.[2] The relationship between their fundamental views on the nature of logical and mathematical truth has, in particular, received very little

[1]This is the paper I presented at *The Bucharest Colloquium for Analytic Philosophy*: *Frege's Philosophy of Mathematics, Bucharest, Romania* in May 2011. As a result of discussions at the colloquium, I significantly reorganized my thoughts on these subjects. This reorganization involved primarily my focus changing more or less exclusively to defending my interpretation of Frege (this was published as Lavers 2013). While there is some overlap between the present paper and the one just mentioned, the heart of the present paper concerns the forces that led to Carnap's position differing from Frege (Sects. 3 and 4) are not at all present in Lavers (2013).

[2]Carnap was a student in three of Frege's lecture courses. He recounts this experience in (Carnap (1963b). Carnap's lecture notes for Frege's courses are now published as Reck and Awodey (2004).

G. Lavers (✉)
Concordia University, Montreal, Canada
e-mail: glavers@gmail.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_15

attention. It is this topic that I wish to address in the present paper. I will show that their goals in this area are actually quite well-aligned.

According to Frege, in putting forward an account of arithmetic we need to respect ordinary usage, but at the same time we are permitted some deviation from ordinary usage. If this is true, then what Frege is doing in the *Foundations of Arithmetic* is giving what Carnap would later (Carnap 1950b) call *an explication*. Much of the paper will be devoted to exploring the differences between Frege and Carnap's views in the foundations of mathematics and logic. Many central claims made by Frege in this area could simply no longer be held by the early 1930s. Carnap, I will argue, can be seen to hold a position on the foundations of logic and mathematics that is a maximally Fregean position in the face of certain technical discoveries.

In the next section, I present an interpretation of Frege according to which he is not attempting to defend realist theses. *Frege's most realist sounding pronouncements are not statements of theses, but statements of desiderata for a definition of number*. In the third section, I look at both Frege's and Carnap's most general theses regarding logical and mathematical truth. Here I show that although Carnap's position is certainly different from Frege's, it is not differences of philosphical temperament, but mathematical developments, that explain this difference. In the fourth section I consider Carnap on ordinary notions. Carnap's explicit rhetoric may lead one to believe that Carnap had no use for, for instance, our ordinary notion of mathematical truth. One might assume that Carnap was interested in constructing arbitrary systems and then assessing their pragmatic value. I argue here that this is not the case. In the final section, I address Tyler Burge's interpretation of Frege. On this interpretation Frege is explicitly presented as a realist in a distinctly unCarnapian way. I argue that the evidence that Burge advances does not show what he intends it to show.

Frege, Realism, and a Definition of Number

Frege thought that arithmetic concerns objects, and that arithmetic truth is independent of any contingent fact concerning human beings. In this sense Frege's position is a realist one. However, I want to claim that Frege's realist pronouncements are not meant as statements of propositions to be defended, but as desiderata of an account of arithmetic. If this is the role realism plays in Frege's philosophy of arithmetic, then it is not the kind of realist position that Carnap would want to dismiss as metaphysical.[3]

[3]Erich Reck, in his (Reck 1997), argues that Frege is not a metaphysical realist in an argument that depends heavily on the use of the context principle. While I don't disagree that the context principle plays an important role, I wish to show that by looking at his criteria for a successful account of number, we can see that Frege's goal is not to establish realist theses.

As just mentioned, I want to claim that Frege's realism is tied to what he thinks a successful account of arithmetic ought to accomplish. It is, therefore of great importance to look closely at what Frege says concerning the conditions under which an account is to be seen as successful. In the introduction to *The Foundations of Arithmetic*, Frege claims:

> Even I agree that definitions prove their worth by being fruitful. (Frege 1884/1980, p. ix)

'Even I' here refers to Frege's strict standards of what is to count as a proof, not to his realist attitudes. Later on in the same work he expands on this thought:

> Definitions show their worth by proving fruitful. […] Let us try, therefore, whether we can derive from our definition of the Number which belongs to the concept *F* any of the well known properties of number. (Frege 1884/1980, §70)

In Section 57, Frege stresses that the goal of providing an account of arithmetic, is to propose one that can play the role required of it in the sciences:

> Now our concern here is to arrive at *a concept of number usable for the purpose of science*; we should not therefore, be deterred by the fact that in the language of everyday life number appears also in attributive constructions. That can always be got round. (Frege 1884/1980, §57, my emphasis)

For Frege, the properties we ordinarily take numbers to have should follow from a definition of number. If we could give an account of number that would be suitable for the role numbers play in mathematics (including applications), then it would seem Frege would view this account as successful. If one examines Frege's criticism of other accounts of arithmetic, we see that he *applies the same standard*. For instance in criticizing psychologistics accounts of arithmetic, Frege argues that the ordinary properties of number cannot be recovered and that this identification of mathematical objects with ideas imports far too much that is foreign into arithmetic. Consider what Frege says in one of his first criticisms of a rival account of arithmetic in the *Grundlagen*:

> When Stricker, for instance, calls our ideas of number motor phenomena and makes them dependent on muscular sensations, no mathematician can recognize his numbers in such stuff or knows what to make of such propositions. (Frege 1884/1980, p. v)

Stricker's definition is to be rejected because it has no similarity with numbers as we ordinarily conceive them.[4] Again consider the following attack on a psychologistic account of number:

> If the number two were an idea, then it would have straight away to be private to me only. Another man's idea is, *ex vi termini*, another idea. We should then have it might be many millions of twos on our hands. We should have to speak of my two and your two, of one two and of all twos. If we accept unconscious ideas, we should have unconscious twos among them, which would return subsequently to consciousness. As new generations of

[4]Both Carnap and Frege maintain that uses of terms by communities of specialists, prior to a systematization, stand in need of clarification. When I speak of 'our ordinary understanding' I meant this to apply equally to such groups of specialists.

> children grew up, new generations of twos would continually be born, and in the course of millennia these might evolve, for all we could tell, to such a pitch that two of them should make five. Yet in spite of all of this, it would still be doubtful whether there existed infinitely many numbers, *as we ordinarily suppose*. 10^{10}, perhaps, might be only an empty symbol, and there might exist no idea at all, in any being whatever, to answer to that name. (Frege 1884/1980, §27, my emphasis)

There are two ways to interpret this passage where Frege is ridiculing psychologistic accounts of number. One way to interpret this is to think Frege's goal is to show that the claim that numbers are ideas *is false* by showing that it leads to absurdities. If this were his goal, then presumably when he introduces numbers as extensions he would need to argue that it is true that numbers are extensions. Of course, no argument is made that numbers really are extensions (we will look at this point in more detail below). The other way to interpret the above passage, strongly suggested by what I have said so far, is to see Frege as arguing that any analysis that identifies numbers with ideas ends up assigning to numbers many properties that we do not ordinarily take them to have (such as necessarily belonging to a particular person). At the same time, such an identification prevents us from showing that numbers do have many of the properties that we ordinarily think we can demonstrate them to have (such as their infinity). On this interpretation the point is not so much that it is ultimately false that numbers are ideas, but that such an identification will not lead to a successful analysis. That is, identifying numbers with ideas will not lead to a concept of number that could serve as the foundation for arithmetic. If this interpretation is correct, then the identification of numbers with something objective, rather than subjective, is motivated by the fact that the identification of numbers with something subjective will lead to a laughably poor definition of number.

Frege, of course, is a realist in the sense that he wants an account that has numbers as objects and he takes mathematical propositions to state objective truths. These realist aspects of Frege's philosophy of arithmetic, however, are *not philosophical conclusions* as much as *statements of desiderata* of an account of arithmetic. Consider what he says about the status of numbers as objects:

> The self-subsistence which I am claiming for number is not to be taken to mean that a number word signifies something when removed from the context of a proposition, but only to preclude the use of such words as predicates or attributes, *which appreciably alters their meaning*. (Frege 1884/1980, §60, my emphasis)

Frege is not claiming here that we should take numbers to be objects because that is simply the way things are. Frege takes it as a desideratum of an account of arithmetical truth that it treat numbers as objects since the use of numerals as singular terms is so pervasive in arithmetic. We have seen so far that agreement with ordinary use is the important criterion by which accounts of number are to be judged. But, and this is very significant, Frege allows an account to introduce some new properties to number. After he defines the numbers as the extension of certain second-order concepts, he writes:

> That this definition is correct will perhaps be hardly evident at first. For do we not think of the extensions of concepts as something quite different from numbers? […] [I]t is not usual

> to speak of a Number as wider or less wide than the extension of a concept; but neither is there anything to prevent us from speaking in this way, if such a case should ever occur. (Frege 1884/1980, §69)

Frege takes his own analysis to be acceptable because the new properties of numbers implied by the account (as a result of numbers being identified with extensions) are seen by him to be sufficiently harmless and not likely to cause any problems.

In *The Logical Foundations of Probability* (Carnap 1950b), Carnap intends to formulate different explications of the concept of probability. He begins, therefore, with a discussion of the criteria a proper philosophical explication should meet. An explication ought to be *simple*, *fruitful*, *precise* and the refined notion should be *similar* to the ordinary one—although perfect overlap is not required. Frege's analysis of number in the *Foundations of Arithmetic* fits Carnap's account of explication, in this sense, perfectly. Frege's account is intended to be simple, precise, and, as he repeatedly insists, fruitful.[5] In addition to this there is great, but not altogether complete, similarity between the ordinary notion and the analyzed notion.

I have been defending a view of Frege's project as seeking to provide an account of arithmetic that could serve as the foundation of mathematics. I have shown that according to Frege's own remarks, the standards by which an account of number is to be judged are whether the properties we ordinarily take the numbers to have are recoverable, and whether the account avoids assigning too many foreign properties to number. But, at the same time, Frege is clear that the statements of his systematic language express truths.[6]

I have argued that Frege does not seek to defend a metaphysical thesis that numbers are objects, and in particular extensions. Must he not then claim that since his reconstructed account of arithmetic is a system of true propositions, the metaphysical statement must be true? To see why Frege need not commit himself to the metaphysical thesis, we need to address Frege's views on analyses. In particular we need to address Frege's views on what is known as the paradox of analysis. According to the paradox of analysis, no account can be both informative and correct. The reasoning behind it is as follows. Suppose A is a notion to be analyzed, and we give an account which says something is A if and only if it is B. If B says the same as A, then the account is uninformative. If not, then it is incorrect.[7] I will not here go into any great detail regarding Frege's views on the paradox of analysis, but some remarks on this subject are clearly necessary. Frege was aware of the paradox

[5]Frege's views on *fruitfulness* are discussed in detail in Tappenden (1995).

[6]Of course, Frege clearly rejected thinking of truth in terms of correspondence. See Ricketts (1986, 1996), and Reck (2007) for further discussion of Frege and the correspondence theory of truth.

[7]Frege's views on analyses have been addressed by Michael Beaney in his Beaney (1996) Chap. 5 and his Beaney (2004). Joan Weiner concentrates on what we must conclude regarding the truth of both sentences of ordinary language and of the systematic language in her Weiner (2007).

of analysis.[8] He was also acutely aware that everyday discourse (and even pre-systematized scientific discourse) contains expressions that lack completely clear meanings:

> Now it may happen that this sign (word) is not altogether new, but has already been used in ordinary discourse or in a scientific treatment that precedes the truly systematic one. As a rule, this usage is too vacillating for pure science. (Frege 1906/1984, pp. 302–3)

Frege gave several proposed solutions to the paradox of analysis.[9] At one point he thought the distinction between sense and reference could solve the paradox. That is, an analysis need preserve reference but not sense. He soon realized, however, that mere reference preservation is far too coarse of a notion to fill this purpose. In his final attempt at a resolution Frege says the following:

> If we have managed in this way to construct a system of mathematics without the need for the sign *A* [a sign with an existing sense], we can leave the matter there; there is no need at all to answer the question concerning the sense in which—whatever it may be—this sign had been used earlier. In this way we court no objections. However it may be felt expedient to use the sign *A* instead of the sign *B* [the newly introduced sign]. But if we do this, we must treat it as an entirely new sign which has no sense prior to the definition. We must explain that the sense in which this sign was used before the new system was constructed is no longer of any concern to us, that its sense is to be understood purely from the constructive definitions that we have given. In constructing the new system we take no account, logically speaking, of any thing in mathematics that existed prior to the new system. Everything has to be made anew from the ground up. Even anything we accomplish by our analytical activities is to be regarded only as preparatory work which does not itself make any appearance in the system itself. (Frege 1914/1979, p. 211)

Here we see Frege saying that we begin by identifying the properties of the ordinary notion. But this is mere preparatory work.[10] The ordinary notion is eventually replaced by the new notion whose meaning is given entirely by its definition and no further relation to the old notion is supposed. The use of the same

[8]See Frege (1894/1984) cited in Chap. 5 of Beaney (1996).

[9]Again see Beaney (1996).

[10]Nelson (2008) gives an interpretation of Frege, and specifically he focuses on Frege's views on analysis, according to which this task of identifying what is true of our ordinary notion is given primary importance. However, Nelson's position is based on a single quotation from Frege (1914/1979) in which Frege discusses the task of examining our ordinary notions. In this work Frege divides the task of *the construction of the system* into two subtasks. There is the merely preparatory subtask that involves examining our ordinary notions to see what is true of them. After this there is the step of providing a constructive definition, in which all ties to the pre-existing sense are severed. The emphasis that Nelson places on the first stage is out of keeping with the importance that Frege himself places on it. In fact, immediately after the quote that Nelson uses—where Frege talks about examining our ordinary notions—Frege says the following: "Work of this kind is very useful; it does not, however, form part of the construction of the system, but must take place before hand. Before the work of the construction is begun, the building stones have to be usable; i.e. the words signs, expressions, which are to be used must have a clear sense, so far as a sense is not to be conferred on them in the system itself by means of a constructive definition." (Frege 1914/1979, p. 211) Frege reserves the term *analysis* for the first task, but it is providing the constructive definition that is the more important task.

term is purely for expedience. Frege's position is that in giving an account of some notion *A* we ought to replace it with a fruitful notion that has some similarity with the old notion, but in general the new notion will be far more precise. The similarities to Carnap's views on explication (mentioned above) are striking.

Of course, it may now be objected that I am reading Frege's 1914 views expressed in 'Logic in Mathematics' back into *Grundlagen*. Obviously this is not my intention at all. I do not mean to claim that when writing *Grundlagen* Frege already had his later account of analysis in the back of his mind. I mention these later views on analysis for two reasons. First, I mention them because of the striking similarities to Carnap's account of analysis. And secondly, I mention the later position because although Frege may not have had his ideas on the paradox of analysis fully worked out by the time of *Grundlagen*, his identification of numbers with extensions is fully compatible with this later view. What his remarks on the identification of numbers with extensions is not compatible with, is the view that in giving an analysis we seek to uncover what has always been true of the notion. That is to say, while his *general* views on what is achieved by an analysis were not yet worked out, what he actually does in *Grundlagen* is an *example* of such an analysis.

Frege and Carnap on the Foundations of Mathematics and Logic

I have argued in the preceding section that Frege is not a concerned to defend realism as a set of metaphysical theses, and that his attitude towards his account of arithmetical truth is in fact much more Carnapian than it is standardly taken to be. It is Carnapian in the sense that Frege is in fact more concerned to articulate a theory of arithmetic that could play the necessary role as the foundations of mathematics. I want to argue now that Carnap's views on logical and mathematical truth are in a sense maximally Fregean. That is, regarding logical and mathematical truth, Carnap attempts to preserve Frege's position as much as possible in the face of certain technical discoveries. I will begin, then, by outlining Frege's fundamental views on the foundations of logic and mathematics. Frege's overall goal in the foundations of mathematics (excluding geometry) was to establish the logicist thesis.[11] We can state this thesis as follows:

> Frege thesis 1:
> Any mathematical truth (excluding geometry) can be given a gapless derivation from basic logical truths (together with definitions).

Frege devoted much of his life's work to establishing this thesis. This stance on the fundamental epistemology of mathematics raises an obvious question. If mathematical truths have as their ultimate justification a derivation from basic logical principles, what is the ultimate justification of basic logical principles? This

[11]The difference in their respective attitudes to geometry will not be discussed in the present paper.

is a matter on which Frege says notoriously little. Frege is most explicit about this question in *The Basic Laws of Arithmetic*, where he writes:

> The question of why and with what right we acknowledge a law of logic to be true, logic can only answer by reducing it to another law of logic. Where that is not possible, logic can give no answer. If we step away from logic, we may say: we are compelled to make judgements by our own nature and by external circumstances; and if we do so, we cannot reject this law — of identity for example; we must acknowledge it unless we wish to reduce our thought to confusion and finally renounce all judgement whatever. I shall neither dispute nor support this view; I shall merely remark that what we have here is not a logical consequence. What is given is not a reason for something's being true, but for our taking it to be true. Not only that: this impossibility of ours of rejecting the law in question hinders us not at all in supposing beings who reject it; where it hinders us is in supposing that these beings are right in so doing, it hinders us in having doubts whether we or they are right. At least this is true of myself. If other persons presume to acknowledge and doubt a law in the same breath, it seems to me an attempt to jump out of one's own skin against which I can do no more than urgently warn them. Anyone who has once acknowledged a law of truth has by the same token acknowledged a law that prescribes the way in which one ought to judge, no matter where, or when, or by whom the judgement is made. (Frege 1893/1967, p. 15)

Frege is careful here to distinguish the logical questions of the ultimate grounds of a principle from the non-logical question about why we hold it to be true. On the first question he clearly asserts that basic logical laws can be given no justification. On the second question he speaks only in a hypothetical and noncommittal manner. Despite this, there are at least three clearly identifiable theses in this passage.

> Frege thesis 2:
> Basic logical truths cannot be justified.
> Frege thesis 3:
> Why one accepts a law of logic is a very different question from why it is true.
> Frege thesis 4:
> There is only one logic (classical) and it is universally valid.

What I wish to do in this section is to examine the relations between this Fregean position on the nature of logical and mathematical truth and that of Carnap. Carnap's major work dealing with logical and mathematical truth is *The Logical Syntax of Language* (*Syntax* hereafter). It will be this work that will be the focus of the present section.[12] In this work Carnap outlines two languages, which he calls Languages I and II, and puts forward his famous Principle of Tolerance:

[12]In his 1931 address on logicism, published as (Carnap 1983), Carnap has a much more optimistic view as far as logicism is concerned. Here he defines the natural numbers as the numerically definite quantifiers and embraces Russell's if-thenism with respect to he axiom of infinity (and Choice). Carnap, here, sees the major obstacle for logicism being a justification of Ramsey's simple type theory that does not involve Ramsey's platonism. Also of interest concerning Carnap's views on the nature of logical and mathematical truth are Carnap (1942, 1939/1955, 1950a, b).

Also, I do not want to make claims about Carnap's conscious motivations. I do not want to say that Carnap held the position he did *in order to* remain maximally Fregean (as I am using the term). Nor, of course, do I want to imply that other thinkers not mentioned (such as Hilbert) had no significant impact on shaping Carnap's philosophical views at this time. I merely wish to argue that Carnap's position is maximally Fregean in the defined sense.

> In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e., his own form of language, as he wishes. All that is asked of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactic rules instead of philosophical arguments. (Carnap 1934/1937, §17)

It seems like we have moved worlds apart from the position of Frege who describes the attitude of the logical pluralist as akin to attempting to jump out of ones own skin. However, as I will argue, their positions are much closer than may be suggested at present.

If thesis 1–4 characterize a Fregean position on the ultimate justification of logic and mathematics, then three technical developments in the early twentieth century (i.e., in the years leading up to *Syntax*) show a Fregean position to be untenable. These developments are Russell's paradox, the discovery of non-classical logic, and Gödel's first incompleteness theorem. I want now to explore the impact of each of these developments on the position of *Syntax*. The position of *Syntax* can be seen as an attempt to preserve the Fregean position to the greatest extent possible given these technical developments.

Let us begin with the discovery of non-classical logics.[13] In the presence of merely hypothetical people who dispute a principle of logic, Frege's answer to the epistemology of logic may seem unsatisfying. But in the presence of actual people who dispute a principle of logic, Frege's position defined by theses 2–4 is simply indefensible. Once there are well worked out alternatives to classical logic (which avoid reducing our thought to confusion), accepting theses 2–4 amounts to holding that classical logic is the only true logic despite the absence of any reason why this is the case. This position is clearly too dogmatic to be defensible. In the presence of alternative logics, unless one insists on holding the dogmatic position just described, there are two options for the Fregean. One could continue to hold 3 and 4 and abandon 2. In this case one continues to hold that classical logic is the one true logic, but attempt to justify classical logic over any of its rivals. While this position remains Fregean in its commitment to classical logic, it would have to introduce an unFregean epistemology of logic.

The tension introduced by the discovery of non-classical logics is between thesis 2 and thesis 4. So, the other option is to accept 2 and 3, but abandon 4. That is, accept that it is impossible justify a law of logic, and keep questions of why we accept a principle completely distinct from questions concerning the truth of the principle, but at the same time allow for many systems of logic to have equal claim to truth. This option maintains the Fregean position that no justification can be given for basic logical laws but relativizes the notion of logical truth to various systems of logic. This is of course the option that Carnap endorses with his famous *principle of tolerance*.

[13]There were of course other bifurcating pressures on logic at the time. Ought one accept type theory at all? If so, ought it be simple or ramified?

Carnap's answer is that prior to accepting a consequence relation which defines a system of logic there is no answer to what is a logical truth. Once, however, a consequence relation is accepted, the logical truths are simply the consequences of the null class. As argued above, in the face of, for instance, Heyting's formalization of intuitionistic logic—which is explicitly mentioned in *Syntax*—it becomes unreasonable to hold Frege's theses 2–4. Carnap's relativism can be seen as an attempt to save two of the three Fregean theses on the foundations of logical knowledge. Carnap certainly maintains that whatever reasons we have for choosing a particular system, are not reasons to suppose that what is provable in that system is ultimately true. Carnap is therefore in clear agreement with 3. Also, while we may be able to show that a claim follows from the primitive sentences in a particular system, there is no question of justifying the primitive sentences. Thus Carnap is clearly in agreement with Frege on 2.

In the presence of alternative logics, one of Frege's theses concerning the basic epistemology of logic must be abandoned. Some drastic change is required. Either one must provide an account of how a system of basic logical laws can be justified over any rival system, or agree with Frege that such justification is impossible but disagree with him concerning the privileged status of the laws of classical logic. Either option involves a radical change in one's epistemology of logic. It might be argued that by taking the first of these two routes one would end up with a position closer to that of Frege's. Nevertheless, Carnap does not stray from Frege more than is necessary. Although abandoning thesis 2 might in principle allow one to stay closer to Frege, there is still no generally accepted defense of the universal validity of classical logic. So that Carnap did not follow this strategy cannot be seen as deviating from Frege more than is necessary.

The second technical development that caused Carnap to deviate from Frege's position is the discovery of Russell's paradox. Frege believed it to be a matter of logic that there are infinitely many objects. Unfortunately the detailed proof of this relied on his Basic Law V, which implies that any property defines an extension. As is now well known, this leads immediately to Russell's paradox. Basic Law V was a principle of complete generality that could be used to demonstrate the existence of infinitely many objects. As a principle of universal generality its claim to be a principle of logic could be defended. Once Basic Law V was shown to be inconsistent, and no other suitably general, yet consistent, principle could take its place, it became apparent that the derivation of arithmetic requires an axiom of infinity.[14]

Given that arithmetic could no longer be derived from *purely logical* axioms, because of the need for an axiom of infinity, Carnap's next move was to follow his own principle of tolerance. Since we can lay down any primitive sentences and rules of inference that we like (or equivalently any consequence relation), we can

[14]I mean here an axiom asserting that there are infinitely many individuals, not a set with infinitely many members.

simply include the required mathematical sentences we need among the primitive truths. For instance in both languages I and II Carnap includes the claims that zero is not a successor and that distinct numbers have distinct successors among the primitive sentences. Carnap believes that by doing so he is not completely abandoning logicism. The reason for this is that in the context of the principle of tolerance, and given that a precise demarcation between logic and mathematics "has so far not been given by anyone" (§84), a reduction of mathematics to logic becomes both unclear and unnecessary. It is unclear given the absence of a clear demarcation. It is unnecessary because in the context of the principle of tolerance, logic and mathematical truths can be seen to have the same status without a reduction of one to the other. Either may be freely chosen in the construction of a language. This way, Carnap can defend the claim that logical and mathematical truths are analytic, without needing to defend the axiom of infinity as a principle of pure logic. By doing so, he achieves one of the main goals of logicism—showing mathematical truths to be analytic—without requiring a reduction of mathematics to logic.[15]

Abandoning any distinction between mathematics and logic may seem like a significant step away from a Fregean position. But consider what Frege says about a demarcation between logic and mathematics in his 1885 piece *On Formal theories of arithmetic*:

> [N]o sharp boundary can be drawn between logic and arithmetic. Considered from a scientific point of view, both constitute a unified science. If we were to allot the most general basic propositions and perhaps their most immediate consequences to logic while we assign their further development to arithmetic, this would be like separating a distinct science of axioms from that of geometry. Of course, the division of the entire field of knowledge into the various sciences is determined not merely by theoretical but also pragmatic considerations; and by the preceding I do not mean to say anything against a certain pragmatic division: only it must not become a schism, as is at present the case to the detriment of all sides concerned. (Frege 1885/1984), p 112–113)

Frege does, in this work, still talk of a reduction of all mathematical forms of reasoning to purely logical forms of reasoning. Carnap, however, in abandoning any distinction between logic and mathematics, and at the same time any attempt to reduce one to the other is not far from Frege's own position. The difference is that at this time Frege still believed that it was possible to demonstrate the existence of infinitely many objects on grounds that could be defended as purely logical. In the wake of Russell's paradox, this began to seem highly unlikely.

We have so far seen that the discovery of non-classical logic pushes Carnap to formulate the principle of tolerance. Russell's paradox, by showing the need for the axiom of infinity in the derivation of mathematics, leads Carnap to abandon the project of reducing mathematics to logic. The other technical discovery that had a major impact on Carnap was of course Gödel's first incompleteness theorem. No longer could it be claimed, as Frege thought, that every truth of arithmetic could be

[15]Weiner (1990) argues that this is the principal motivation of Frege's logicism.

given a gapless deduction from basic logical laws. Gödel's incompleteness theorems actually motivated Carnap to write *Syntax*.[16] Carnap's response to Gödel's theorem was to distinguish between derivability and consequence. In Language I, Carnap adds an ω-rule when defining the notion of consequence. In Language II Carnap begins by defining 'analytic in Language II' and then defines consequence in terms of it. Here he gives what is actually a semantic definition of 'analytic in Language II'.[17] While not every arithmetical truth expressible in the system will be derivable from the basic sentences, all such sentences are consequences of the basic sentences.

Given his formulation of the notion of consequence, we can now see that Carnap holds a modified version of Frege's first thesis (discussed above). Recall that for Frege the thesis was:

> Any mathematical truth (excluding geometry) can be given a gapless derivation from basic logical truths (together with definitions).

Carnap's significant reformulation of this thesis is:

> Any mathematical truth (including pure geometry) is a consequence of the primitive logico-mathematical sentences in a suitable system.

This reformulation involves all three of the changes Carnap made in response to the technical results discussed above. We saw that the discovery of non-classical logic, Russell's paradox, and Gödel's first incompleteness theorem lead respectively to the principle of tolerance, the abandonment of a sharp division between logic and mathematical truths (and any attempt at reduction), and the formulation of the notion of consequence. The principle of tolerance motivates the relativization to suitable systems. This, as we saw, was occasioned by the development of non-classical logics. Another difference between the two principles is that no longer is mathematics said to follow from pure logic, but follows from logico-mathematical primitive sentences. We saw that this is a result of Russell's paradox which made evident the need for an axiom of infinity. And the final difference between the two principles is that the notion of derivability is replaced with a notion of consequence. Although Carnap's final position is quite different from Frege's, Carnap can be seen as attempting to remain as close to the Fregean position as he can. It is not that Carnap strayed from the path that Frege set out on, Carnap attempted to stay on course, as much as possible, despite some major obstacles.

[16]See Carnap (1963b).

[17]Of course, Carnap calls the definition syntactic. However, Carnap uses the term 'syntax' at this time to include much of what we now call semantics (see Creath 1990).

Carnap and Ordinary Notions

One might think, given what has been said so far, that Frege and Carnap are still quite different in one respect. One might accept that Frege was not trying to defend a metaphysical thesis, but merely wanted to provide an account of arithmetic that preserves the properties that we take our ordinary notions of number to have. But, one could suppose, this itself is a difference between the two positions. That is, Frege was interested in defining a notion of arithmetical truth that is in keeping with our ordinary notions, while Carnap was interested in defending the construction of a 'boundless ocean of unlimited possibilities' of possible systems.[18] I will argue in this section that this is not the case. As a matter of fact, our ordinary notions have, in a sense, a more important role to play in Carnap's philosophy than they do in Frege's.

As I have argued elsewhere, Lavers (2008), Carnap is concerned in *Syntax* with preserving our ordinary notion of mathematical truth.[19] Language II was meant to include all of classical mathematics. It was not meant as a substitute for classical mathematics which would be sufficiently suitable for practical purposes. All the truths of classical mathematics turn out as analytic in Language II. That Carnap means to capture our ordinary notion of mathematical truth, in defining 'analytic' for Language II, can be seen in the following passage:

> As a result of Gödel's researches it is certain, for instance, that for every arithmetical system there are numerical properties that are not definable, or, in other words, indefinable real numbers […]. *Obviously it would not be consistent with the concept of validity of classical mathematics* if we were to call the sentence: "All real numbers have the property M" an analytic sentence, when a real number can be stated (not, certainly, in the linguistic system concerned, but in a richer system) which does not possess this property. (Carnap 1934/1937, §34C, my emphasis)

Here we see Carnap clearly intends his 'analytic in Language II' to capture our ordinary notion of truth of classical mathematics. Carnap explains that if we interpret sentences that quantify over all properties as quantifying merely over the definable properties, then we get something that does not fit with our notion of classical mathematical truth. The definition provided involves assigning an intended domain to every logical type in the language. This is not, of course, a purely formal definition. As pointed out in the previous section, in response to Gödel's first incompleteness theorem, Carnap distinguishes between the formal notion of derivability, and the more liberalized notion of consequence. Matters are complicated somewhat, however, by Carnap's use of the term 'formal' at the time of *Syntax*. There Carnap uses the term 'formal' to mean *not concerned with meaning*. This obviously has little connection with the term 'formal' as we standardly use it today. On the basis of his interpretation of 'formal' he declares that all his various

[18]This phrase is taken from the forward to *Syntax* (pp. xv).

[19]Since I have discussed Carnap on ordinary notions in detail elsewhere, this discussion will be kept relatively brief.

definitions provided in *Syntax* are 'formal'.[20] However, he is clear to point out that his definition of 'analytic in Language II' is 'indefinite' (the analytic truths form non recursively enumerable set).

In the Schilpp volume on Carnap, there is an exchange between Carnap and E. W. Beth that is particularly telling for the present purposes. Here Beth complains that some of Carnap's definitions, that of 'analytic for Language II' in particular, allow for non-standard interpretations. On these non-standard interpretations certain mathematical falsehoods come out as analytic and certain truths come out contradictory. In response to this Carnap writes:

> I certainly agree with Beth when he says (in §5): "We find in Logical Syntax also concepts which, though defined in a purely formal way, are clearly inspired by a non-formal interpretation which, if made manifest, would imply a return to Frege's logicism." And perhaps I would also agree with his further statement: "I think that it is even possible to show that, in the absence of such a non-formal, intuitive interpretation, the whole edifice of Logical Syntax would miss its purpose". (Carnap 1963a, pp. 928)

Here we see, first of all, Carnap happily accepting the comparison to Frege. Beth intends "Frege's logicism" to connote a metaphysical position to be avoided, but Carnap does not see the comparison to Frege in this way. We also see Carnap conceding that our ordinary understanding of mathematical concepts plays a role in the construction of his languages. When spelling out the definition of 'analytic in Language II', Carnap says for instance that a sentence that quantifies over all properties of the natural numbers, for instance $(\forall F)\mathbf{M}F$, is to be evaluated in terms of all possible valuations. That is, for it to be true, any arbitrary property of the natural numbers must satisfy **M**, not just the definable ones. Carnap dismisses Beth's non-standard interpretations of 'all possible valuations', since the non-standard interpretations are clearly not what is meant by the phrase 'all possible valuations'. Saying that without these intuitive interpretations *Syntax* would "miss its purpose" shows that (at least in retrospect) Carnap saw one of the goals of *Syntax* to be to capture our ordinary notions.

Above we looked at Frege's ideas on the paradox of analysis. We saw that for Frege our ordinary notions play only a preliminary role in an analysis. Once we have identified the properties we wish to preserve, the ordinary notion is replaced with a more precise notion. The properties of the precise notion are to follow formally from its definition. Once the definitions have been given, there is no further role for the ordinary notion to play. Given Gödel's first incompleteness theorem, it becomes clear that we cannot capture all of the properties of most interesting mathematical notions by purely formal definitions. In defining 'analytic in Language II' Carnap evaluates claims about all natural numbers in terms of the set $\{0, 0', 0'' \ldots\}$, and he evaluate claims about all properties of the natural numbers in terms of all subsets of the natural numbers—not just all definable properties. We require our ordinary understanding in order to both interpret the three dots in the above set, and to interpret the phrase 'all subsets'. Our ordinary notions, therefore,

[20] I discuss Carnap's use of the term 'formal' in both Lavers (2004, 2008).

are required in order to understand the definitions. No longer can the ordinary notion be completely cast aside, as Frege supposed, once a precise definition is given. So, we see that our ordinary notions have, in this sense, a more important role in Carnap's philosophy of mathematics than they do in Frege's.[21]

Burge on Frege

In this final section I wish to discuss the views of a commentator of who gives an interpretation of Frege at odds with the interpretation presented here. Tyler Burge has defended a view of Frege as a Euclidean rationalist. Frege's goal on this reading is to derive all of arithmetic from self evident axioms. What distinguishes Frege from more traditional Euclidean rationalists, on Burge's view, is Frege's fallibilism concerning such a project. So far, there is nothing that is clearly at odds with the interpretation of Frege presented here. Burge, however, interprets Frege as holding a very strong form of realism. This realism is to be understood ontologically, and Burge warns against any "non-metaphisical" reading of Frege's platonism. The contrast with the view defended here should now be apparent. In fact Burge writes:

> I would not take very seriously a reading of Frege as a Carnapian. […] I think it clear […] that Frege was trying to provide a rational foundation for mathematics—in a way Carnap would have regarded as misguided. Frege saw reason, not practical recommendation, as giving us logical objects (e.g. *FA* §105). There is nothing remotely akin to Carnap's Principle of Tolerance either in Frege's philosophical pronouncements, or even more emphatically, in his temperament. (Burge 2005, Chap. 8, p. 304)

The first thing that can be said about this contrast between Frege and Carnap is that, in several respects, it is entirely correct. As we saw above, Frege was not a pluralist concerning logical or mathematical truth. The principle of tolerance was one of the three main differences identified between Frege and Carnap. Another difference was that, at least before Russell's paradox, Frege believed that the existence of infinitely many objects could be proved from principles of pure logic. So, Burge's point about reason giving us logical objects for Frege cannot be rejected.

The main difference between the interpretation presented in this paper, and Burge's interpretation, then, is the claim that Frege was trying to provide a rational foundation for mathematics in a distinctly unCarnapian way. Burge believes it is important to interpret Frege as a realist of a clearly metaphysical kind. Despite this he says "Most of Frege's uses of his metaphysical view are defensive. His metaphysical remarks ward off idealist, physicalistic, psychologistic, reductive or deflationary positions because he thinks that they prevent a clear understanding of

[21]Interestingly, Carnap agrees with Frege in holding that our ordinary notions may not be completely precise. In his response to Beth, Carnap claims that our ordinary notions may not be 'univocal'.

the fundamental notions of logic and arithmetic." (Burge 2005, Chap. 8, p. 302)[22] It was argued above that Frege's most metaphysical sounding pronouncements were tied to his desiderata for a foundation for arithmetic, and were not the expression of a philosophical position about the ultimate nature of reality. What I wish to argue now is that Frege's realism does not need to be interpreted in a manner that Carnap would clearly wish to reject.

It was argued above that Frege was largely unconcerned with metaphysical questions of ontology. When he addresses the question of whether numbers are objects, he takes a purely pragmatic stance. The same is true when it comes to considering the truth-values (as objects):

> How much simpler and sharper everything becomes by the introduction of truth-values, only detailed acquaintance with this book can show. These advantages alone put a great weight in the balance in favor of my own conception, which indeed may seem strange at first sight. (Frege 1893/1967, p. 7)

Similarly, concerning course-of-values, Frege states, "The introduction of courses-of-values of functions is a vital advance, thanks to which we gain far greater flexibility." (Frege 1893/1967, p. 6) Burge does not deny that Frege gave pragmatic reasons for holding certain principles. And we see this is true of exactly the principles which Frege uses to introduce a new kind of object. However, Burge rarely uses "pragmatic" without scare quotes. The reason for this is given in a footnote where he writes: "I hope that it is clear that by calling epistemic considerations "pragmatic" I am in no way implying that Frege thought them any less able to put us on to truths about a reality independent of our practice." (Burge 2005, Chap. 9, p. 341)

Burge puts a great deal of emphasis on the point that the truth of mathematical claims for Frege is *indepedent of our practices*. This is where, according to Burge's interpretation, the primary difference between Frege and Carnap lies.[23] There are also warnings about understanding Frege in a deflationist way. But to identify what exactly Burge means by this and to what extent this term can be applied to Carnap (and Frege) would take us too far off course. These warnings are, however, closely related to his insistence that the truth conditions for mathematical claims are *independent of us*. I will therefore focus on the question of the independence of truths about abstract objects on us and our practices.

The first thing to be noticed, when asking if a truth is independent of us and our practices, is that there are several senses in which this might be true. There are at least three distinct ways that truth conditions can be *dependent or independent of us and our practices*

[22] I am not aware of any evidence that Burge puts forward with respect to the claim that Frege is trying to ward off deflationary positions.

[23] Burge avoids going into a detailed comparison with Carnap, but takes it to be clear that there are obvious differences between their positions.

1. We may or may not fix the truth conditions for a class of sentences by something like pragmatic choice.
2. We (or our practices) may or may not figure in those truth conditions.
3. Whether these truth conditions obtain is some kind of basic metaphysical fact about the world.

There is clear evidence that Frege thought mathematical truths are independent of the thoughts or activities of any human. But what Burges needs, to maintain that Frege meant this in a distinctly unCarnapian sense, is that Frege wanted to claim their independence in specifically the third of the above senses. Carnap would clearly dismiss questions of type 3 as metaphysical psuedoquestions. As we saw above though, Frege was unconcerned with such questions as whether numbers are really objects. Also, of course, he rejects any explanation of truth in terms of correspondence.

What I now wish to show is that in terms of the other senses, Frege and Carnap's (post semantic) views coincide.[24] I do not see anything in Frege that would amount to holding that we cannot fix the sense of a certain class of sentences by something like pragmatic choice. In fact, he explicitly states otherwise:

> Since it is only in the context of a proposition that words have meaning, our problem becomes this: To define the sense of a proposition in which number words occur. That, obviously, leaves us still a wide choice. (Frege 1884/1980, §62)[25]

So Frege does not insist on independence of type 1. Notice that this type of (in) dependence concerns only the sentences and not the propositions expressed by those sentences. Saying that we (and our pragmatic choices) may play some role in determining the truth condition of sentences should not be seen as a very radical position. That both Carnap and Frege would hold that mathematical sentences may be dependent on us in the sense of (1) should be acceptable to even those who see Frege committed to defending metaphysical theses. What Frege says, however, commits him to mathematical truths being independent in senses (2) Does this amount to an unCarnapian metaphysical position? Certainly not. Frege held that sentences express thoughts, and that these thoughts can hold or fail to hold in a way completely independently of us. As we will see below, Carnap says the same thing concerning propositions.

In 'The Thought' (Frege 1918/1956) Frege clearly claims that thoughts are independent of us in what is often seen as a strongly metaphysical sense. He even, notoriously, goes so far as to claim that they exist in a 'third realm'. It may seem that these remarks clearly commit him to a metaphysical position that Carnap could not have accepted. This, however, would be a mistake. When Frege speaks of the

[24]Matters of ontology are treated in a quite confused manner in *Syntax*. For an argument to this effect see Lavers (2004). It is for this reason that I focus in this section on Carnap's post semantic views on matters of ontology.

[25]Austin translates 'Satzes' as 'propositions' rather than 'sentences'. However, 'sentences' is a better choice here. Number *words* do not appear in propositions.

third realm it is to contrast thoughts with both objects of the senses and psychological entities. Frege takes it as clear that thoughts are not of either of these types. But this position, that thoughts are non-mental and non-physical, need not be taken in a strongly unCarnapian metaphysical sense. In fact, consider this quote from *Meaning and Necessity* where Carnap discusses propositions:

> We take as the extension of the sentence its truth-value, and as its intension the proposition expressed by it. This is in accord with the identity conditions for extensions and for intensions stated in the preceding section. Propositions are here regarded as objective, nonmental, extra-linguistic entities. It is shown that this conception is applicable also in the case of false sentences. (Carnap 1947/1956, p. 25)

Here Carnap makes a remarkably Fregean move in holding that we may take propositions to be objective, abstract objects since the truth conditions for identity statement involving them have been clearly defined.

Or again consider Carnap on Propositions in 'Empiricism, Semantics and Ontology':

> For example, are propositions mental events (as in Russell's theory)? A look at the rules shows that they are not, because otherwise existential statements would be of the form: "If the mental state of the person in question fulfils such and such a condition, then there is a *p* such that …". The fact that no references to mental conditions occur in existential statements […] shows that propositions are not mental entities (Carnap 1950a, p. 210 (1956))

We see in both of these quotes Carnap dismissing a psychologistic view of propositions. In fact, in the second quote, he argues against psychologism on the grounds that mention of our mental states does not figure in the truth conditions for existential claims regarding propositions. He goes on to say that no mention of people at all appear in the truth conditions for propositions. Carnap therefore holds that propositions are non-psychological, objective entities whose existence is independent of us. So, both Carnap and Frege hold that the truth of claims are dependent on us in sense 1, but independent of us in sense 2. Finally, neither Frege nor Carnap thought it important to address the independence of mathematical truths in the sense of 3 above. There is, therefore, contra Burge, no reason to view the independent truth of certain claims as a metaphysical view that separates Frege from Carnap.

Before concluding, there may be one last objection to the argument presented in this section. It may be argued that, perhaps, Carnap does consider propositions to be objective, non-linguistic, and non-psychological entities, but all this is relative to a framework. We may have some frameworks that include these objective, mind independent propositions, and some frameworks that do not. Surely, the objection continues, this is a significant difference between Frege and Carnap. And of course this is right. Frege was not a pluralist and Carnap was. But rather than standing in conflict with the position defended in this paper it is clearly in complete accord with it. Carnap's acceptance of the principle of tolerance was identified as a principal difference between their positions on the nature of logical and mathematical truth.

Conclusion

I hope to have shown that there are many more interesting connections between Frege and Carnap, on the foundations of logic and mathematics, than the standard picture of Frege as the metaphysician and Carnap the anti-metaphysical philosopher would suggest. When Frege seems most like a metaphysical realist—when insisting that numbers are objects, or when criticizing psychologistic accounts of number—, he is actually making claims about what is expected of an analysis of number. Furthermore, what Frege says about assessing the success of an account of arithmetic sounds very Carnapian. We saw, in addition to this, that Frege's explicit account of how analysis should function is nearly identical to that presented in (Carnap 1950b, p. 210 (1956)).

I argued that while there are clear differences between their positions on the foundations of logic and mathematics, these differences can be traced to a handful of mathematical results in the early twentieth century. The development of non-classical logic led to Carnap's logical pluralism. Russell's paradox led Carnap to abandon the goal of reducing mathematics to pure logic. And finally, Gödel's first incompleteness theorem led Carnap to replace the philosophical role played by gapless derivation with that of the more liberal notion of consequence. I then attempted to clear up a common misconception concerning Carnap and ordinary notions. I argued that Carnap was concerned to preserve ordinary notions. In fact, it was argued that, in a sense, these ordinary notions play a more important role in Carnap's philosophy than they did in Frege's.

Finally, I showed, contra Burge, that Frege's remarks about the independence of mathematical truth on humans and their activities does not point to an important difference between Frege and Carnap. Ultimately, understanding Frege and Carnap as dealing with the same issues and arriving at remarkably similar positions, differing mainly because of what was known at the time, allows for, I hope, a greater appreciation of each of their positions.

References

Beaney, M. (1996). *Frege: Making sense*. London: Duckworth.

Beaney, M. (2004). Carnap's conception of explication: from Frege to Husserl? In S. Awodey & C. Klein (Eds.), *Carnap brought home: The view from Jena* (pp. 117–150). Chicago, La Salle: Open Court.

Burge, T. (2005). *Truth thought reason: Essays on Frege*. Oxford, New York: Oxford University Press.

Carnap, R. (1934/1937). *The logical syntax of language*. London: Routledge & Kegan Paul.

Carnap, R. (1939/1955). Foundations of logic and mathematics. In O. Neurath, R. Carnap, & C. Morris (Eds.) *International encyclopaedia of unified science*. Chicago: University of Chicago Press, combined ed.

Carnap, R. (1942). *Introduction to semantics*. Cambridge, MA: Harvard University Press.

Carnap, R. (1947/1956). *Meaning and necessity: A study in semantics and modal logic* (2nd ed.). Chicago, London: University of Chicago Press.
Carnap, R. (1950a). Empiricism, semantics and ontology. *Revue International de Philosophie*, *4*, 20–40. (Reprinted in *Meaning and necessity: A study in semantics and modal logic* (2nd ed.). Chicago: University of Chicago Press, 1956).
Carnap, R. (1950b). *Logical foundations of probability*. Chicago, IL: University of Chicago Press.
Carnap, R. (1963a). E. W. Beth on constructed language systems. In P.A. Schilpp (Ed.) *The philosophy of Rudolf Carnap*, vol. XI of *library of living philosophers*, (pp. 927–932). La Salle, IL: Open Court.
Carnap, R. (1963b). Intellectual autobiography. In P.A. Schilpp (Ed.) *The philosophy of Rudolf Carnap*, vol. XI of *Library of living philosophers*, (pp. 927–932). La Salle, IL: Open Court.
Carnap, R. (1983). The logicist foundations of mathematics. In P.B.H. Putnam (Ed.) *Philosophy of mathematics* 2nd ed., (pp. 41–51). Cambridge: Cambridge University Press.
Creath, R. (1990). The unimportance of semantics. In M.F.A. Fine, & L. Wessels (Eds.) *Proceedings of the 1990 Biennial meeting of the philosophy of science association*, vol. 2. East Lansing.
Frege, G. (1884/1980). *The foundations of arithmetic*. Evanston, IL: Northwstern University Press, second revised edition ed.
Frege, G. (1885/1984). *On formal Theories of arithmetic*. In Collected papers on mathematics, logic and philosophy. Basil Blackwell
Frege, G. (1893/1967). *The basic laws of arithmetic: Exposition of the system* (M. Furth, Trans.). Berkeley, Los Angeles: University of California Press.
Frege, G. (1894/1984). Review of E. G. Husserl, philosophie der arithmetik i. In *Collected papers on mathematics, logic and philosophy*. Basil Blackwell.
Frege, G. (1906/1984). On the foundations of geometry: Second series. In B. McGuinness (Ed.) *Collected papers on mathematics, logic and philosophy*, (pp. 293–340). Oxford: Basil Blackwell.
Frege, G. (1914/1979). Logic in mathematics. In F.K.H. Hermes, & F. Kambartel (Ed.) *Poshumous writings*, (pp. 203–250). Oxford: Basil Blackwell.
Frege, G. (1918/1956). The thought: A logical inquiry. *Mind*, *65*(259), 289–311.
Lavers, G. (2004). Carnap, semantics and ontology. *Erkenntnis, 60*(3), 295–316.
Lavers, G. (2008). Carnap, formalism, and informal rigour. *Philosophia Mathematica, 16*(1), 4–24.
Lavers, G. (2013). Frege, Carnap, and explication: 'Our concern here is to arrive at a concept of number usable for the purpose of science'. *History and Philosophy of Logic*, *34*(3), 225–241.
Nelson, M. (2008). Frege and the paradox of analysis. *Philosophical Studies, 137*(2), 159–181.
Reck, E. (1997). Frege's infuence on Wittgenstein: Reversing metaphysics versus the context principle. In W.W. Tait (Ed.) *Early analytic philosophy: Frege, Russell, Wittgenstein*, (pp. 123–85). Chicago, La Salle: Open Court.
Reck, E. H. (2007). Frege on truth, judgement, and objectivity. *Grazer Philosophische Studien, 75* (1), 149–173.
Reck, E. H., & Awodey, S. (Eds.). (2004). *Frege's lectures on logic: Carnap's student notes, 1910–1914*. Chicago, La Salle: Open Court.
Ricketts, T. (1986). Objectivity and objecthood: Frege's metaphysics of judgement. In L. Haaparanta & J. Hintikka (Eds.), *Frege synthesized* (pp. 65–95). Dordrecht: Reidel.
Ricketts, T. (1996). Frege on logic and truth. *Proceedings of the Aristotelian Society Supplement, 70*, 121–140.
Tappenden, J. (1995). Extending knowledge and 'fruitful concepts': Fregean themes on the foundations of mathematics. *Nous, 29*(4), 427–467.
Weiner, J. (1990). *Frege in perspective*. Ithaca, NY: Cornell University Press.
Weiner, J. (2007). What's in a numeral? Frege's answer. *Mind, 116*(463), 677–716.

Author Biography

Gregory Lavers is an Associate Professor at Concordia University in Montreal, where he has been teaching philosophy since 2005. His PhD, which focused principally on Carnap's *The Logical Syntax of Language,* was completed in 2004 at the University of Western Ontario under the supervision of William Demopoulos. His research focuses on issues at the intersection of the philosophy of mathematics, the philosophy of language, and the history of analytic philosophy. Dr. Lavers has published in *Philosophical Studies*, *History and Philosophy of Logic*, *The Review of Symbolic Logic*, *Philosophia Mathematica*, *Erkenntnis*, and elsewhere. He is the co-founder of the Montreal Inter-University Workshop on the Philosophy of Mathematics. He has been awarded grants by FQRSC both as an individual researcher (nouveaux chercheurs program) and as part of a group (subvention de soutien aux équipes).

On the Interconnections Between Carnap, Kuhn, and Structuralist Philosophy of Science

Thomas Meier

Introduction

This work is a historical investigation of the methodological and conceptual connections that exist between Carnap's early work in the *Aufbau*, Kuhn's conception of theory change in empirical science and the so-called *structuralist view of scientific theories* (see Sneed 1971 and Stegmüller 1973, 1976, among others). The connection between the structuralist program and Kuhn's philosophy of science is better known than that between Carnap's early work and the structuralist view. At first sight there might not be an obvious connection between Carnap's early work and the structuralist program, and I argue here that the program of the *Aufbau* should be understood as one mayor influence on Stegmüller and other adherents of the structuralist program. A further source of major influence on the development of the structuralist philosophy of science, I argue, is Kuhn's work.

The two main theses I argue for are therefore the following. First, Carnap's program of knowledge description from the *Aufbau* is analogous with the structuralist program of knowledge description of scientific theories, in at least two senses. First, from a methodological point of view, it emphasizes relational descriptions of our knowledge domains. Second, both Carnap and the adherents of the structuralist program declare themselves to be objective and neutral in questions of realism and antirealism, both generally and more specific in the philosophy of science. My second argument is that Kuhn's ideas find their precise formal reformulation in the framework of the structuralist approach.

T. Meier (✉)
Ludwig Maximilians University Munich, Munich, Germany
e-mail: thomas.meier@lrz.uni-muenchen.de; meier060782@googlemail.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_16

Carnap's Structuralism in the *Aufbau* and the Structuralist View

There exist many interpretations of Carnap's *Aufbau*. Clearly, Richardson's (1998) reconstruction stands out for its rigor in clarity and exhaustiveness. Because I focus on Carnap's structuralist epistemology of the *Aufbau*, I argue that the *Aufbau* is, generally, a program for logically reconstructing our knowledge of the world in structuralist terms. This means that the descriptions of our knowledge are purely structural definite descriptions, as explained in §16. There, Carnap introduces his purely structural definite descriptions of knowledge. Following him, only such descriptions guarantee objectivity:

> Carnap (1928, §16): Every scientific statement can in principle be so transformed that it is only a structural statement. But this transformation is not only possible, but required. For science wants to speak about the objective; however, everything that does not belong to the structure but to the material, everything that is ostended concretely, is in the end subjective. From the point of view of constitutional theory this state of affairs is to be expressed in the following way. The series of experiences is different for each subject. If we aim, in spite of this, at agreement in the names given for the objects constituted on the basis of the experiences, then this cannot occur through reference to the completely diverging material but only through the formal indicators of the object-structures.

The principal aim of Carnap's method is, as he says, to provide an objective way of describing our knowledge of the world. As becomes clearer in his famous *railway example*, only what can be described in relational terms can be known objectively. All that is not described in purely relational terms is, in the end, subjective and should not be the subject matter of scientific inquiry. For my purpose, I now focus on a connection between Carnap's method and the structuralist program. This is a connection that has so far been left implicit. However, I argue that this connection indeed constitutes an essential part of the methodological motivation of the structuralist program. There is only sparse textual evidence of structuralists referring directly to the *Aufbau*. Interestingly, two main figures of the structuralist view, Wolfgang Stegmüller and C. Ulises Moulines, have both written monographs on Carnap's early epistemological program. Stegmüller's *Der Phänomenalismus und seine Schwierigkeiten* (1969) and Moulines' *La estructura del mundo sensible* (1973) provide a detailed analysis of Carnap's work. This reveals at least that some main figures of the structuralist view had a strong interest in Carnap's *Aufbau*. If we consider the date of publication of these works, it is clear that the structuralist program had been developed later, mainly between 1971 with Sneed's *The Logical Structure of Mathematical Physics* and the close of the first phase of consolidation with Balzer, Moulines, and Sneed's 1987 *An Architectonic for Science*. As one main figure of the structuralist program, Moulines expresses the following regarding the *Aufbau*:

> To be more precise, the use of Carnap's *Aufbau* I propose … consists in reinterpreting Carnap's "Konstitutionstheorie" as a formal explication of the notion of an ideal observer, i.e. an epistemic subject provided with the essential constituents of an ideal "observational language" to check any empirical statement made in theoretical science (Moulines 1991, 265).

In structuralist philosophy of science, the primary aim of inquiry is, of course' a different one. Structuralist philosophy of science aims mainly to provide a program for analyzing science by means of a detailed logical analysis of scientific theories. In addition to the logical reconstruction of theories, structuralism also aims to describe the social phenomena of scientific enterprise by incorporating many of Kuhn's original interests (see Balzer et al. 1987). In the framework of the structuralist program, at no point can we find the postulation of an ideal observer in the sense of Carnap. However, in the structuralist program there also exists the motivation of applying logical tools in order to reconstruct our knowledge of the world in purely structural (i.e., relational) terms. In Sect. 3, I outline the structuralist program in detail. There I also show how it further relates to Carnap's *Aufbau* in a methodological sense. Let us now focus on Carnap's structuralism in the *Aufbau*. In §66 he is more concrete on the background motivation and the aim of pursuing an epistemic structuralism:

> How should science come to objectively valid statements, if all its objects are constituted by an individual subject?. .. the solution to this problem lies in that of course the material of the individual streams of experience is completely diverging. .. but certain structural features agree of all streams of experience. Science has to restrict itself to such structural properties, because it aims to be objective. And it can restrict to structural properties, as we have seen earlier, for all the objects of knowledge are not content, but form, and they can be represented as structural entities.

Following Friedman's interpretation, I understand the *Aufbau* to be an outline of one version of an epistemological program of early logical positivism. In the *Aufbau*, as in the structuralist view, there is an emphasis on structural descriptions of our knowledge of the world. In structuralism, this description is made up of our empirical theories, and in Carnap's program, structural descriptions are provided directly from our knowledge. Still, both share the view that our knowledge should be best described in the form of structures. When Carnap starts with structurally describing our knowledge of the world, structuralism describes this knowledge indirectly through the structural description of our empirical theories. The main difference is that Carnap starts at the lowest level, i.e., the level of an ideally constructed epistemic agent, when in the structuralist view the starting level is that of scientific theories.

The Structuralist View of Theories

Sneed's (1971) *The Logical Structure of Mathematical Physics* is the pioneer work of the structuralist view. The methodological tools of this view are developed there and applied to mathematical physics. Although Sneed's original proposal is mainly

motivated in providing a logical framework in a non-statement-view for the description of the logical structure of physical theory, in Chap. 7 he mentions the field of theory dynamics. This part is what relates to Kuhn's philosophy of science. Stegmüller's *Theorienstrukturen und Theoriendynamik* (1973) can be seen together with Sneed's (1971) as the main work in the early development of the structuralist view. Stegmüller offers a detailed analysis of Kuhn's concept of paradigm. Moulines (2008, 163) explains the methodological motivation of the structuralist view as follows:

> Structuralism owes his name to the fundamental thought that the most adequate way of interpreting and understanding what a scientific theory is does not consist in conceiving it as a set of statements, but rather in conceiving it as a form or collection of different types of complex structures, which themselves are built up of simpler structures.[1]

Let us now take a more detailed look at the methodological similarities between Carnap and the structuralist view. When Carnap gives purely structural definite descriptions, he does so with a methodological tool that can be understood to be a *proto-version* of modern set theory. Carnap himself refers to his tool as type-theory by mentioning Russell and Whitehead. The point here is that, at the time Carnap wrote the Aufbau, set theory did not exist in the same form that it did when the structuralist view was developed. However, Russell–Whitehead type-theory was a predecessor of modern set theory.

In the structuralist philosophy of science, the main methodological tool is naïve set theory. In the formal framework of structuralism, an empirical theory consists of its models, which are sequences of the form $\langle D_1,\ldots,D_m; R_1,\ldots,R_n\rangle$. The D_i are so-called basic sets and the R_j are relations constructed on these sets. The elements of the D_i comprise the ontology of the theory, i.e., they contain the "objects" of which the theory is about. The R_j are usually functions mapping empirical objects into real numbers, or some other mathematical entities. The direct methodological predecessor and *founding father* of this method is Suppes (1957). Following Suppes, and the structuralists, a set-theoretic predicate P specifies the following: the type of a structure $\langle D_1, \ldots, D_k; R_1, \ldots, R_n\rangle$ in determining the number k of base sets and the number n of relations; the typification of the relations $R_1,\ldots, R_n$; the axioms that the relations $R_1,\ldots, R_n$ need to satisfy the structure $\langle D_1, \ldots, D_k; R_1, \ldots, R_n\rangle$ to be an instance of the set-theoretic concept P. The existence of the entities which are taken to be the elements of the basic domains of our structures is merely a working proposition. It should also be noted that in the structuralist view, objects only exist within these domains, which are constitutive parts of a structure and specified by the R_j, which are usually functions. More specifically, in the structuralist program, a theory is understood to consist of the following sets of models:

1. *Core of a theory*: An empirical theory T consists of its core K and its intended applications I. K itself consists of sets of potential models Mp, partial potential models Mpp, actual models M, global constraints GC, and the global links GL.

[1]Translated from the German original by the author.

2. *Intended applications*: I are the sets of the intended applications of a theory. These are not formally characterized. Their determination depends on pragmatic constraints.
3. *Theory-elements*: In the structuralist view, empirical theories are not just isolated entities, but instead complex structures with relations to other theories and, within one theory, there are several levels of theory-specialization. This is why theories are usually understood as so-called *theory-elements*. Such an entity is then, formally, the tuple $T = \langle M_p, M_{pp}, M, GC, GL, I \rangle$.
4. *Potential models*: A set of potential models (Mp) fixes the general framework, in which an actual model of a theory is characterized. All entities that can be subsumed under the same conceptual framework of a given theory are members of the sets of the potential models of this theory. Sets of partial potential models (Mpp) represent the framework of data for the corroboration or refutation of the theory in question. The concepts in Mpp can be determined independently of T. Terms which are theoretical (and proper to T) in the potential models of the respective theory are cut out. Sets of models that not only belong to the same conceptual framework but also satisfy the laws of the same theory are called the sets of actual models (M) of a theory T.
5. *Global constraints*: Furthermore, it is a fact that local applications of a scientific theory may overlap in space and time. For this purpose, the formal notion of *global constraint* is introduced. The sets of bf global constraints (GC) are formal requirements that restrict the components of a model in dependence of other components of other models. Constraints express physical or real connections between different applications of a theory, i.e., the inner-theoretical relations. To explain it intuitively, think of a physical object that is part of a system. This object, say, a certain train wagon, must have the same weight, no matter to which physical system that wagon belongs. It may stand on a railroad somewhere in Nebraska, or on a railroad close to Berlin. The same wagon has the same weight if we think of physical systems on Earth. This fact cannot be overlooked. Because of such overlaps, the notion of constraint is required in the structuralist view.
6. *Global links*: Some empirical theories deal with the same or very similar domains of objects. Particles have the same mass in classical collision mechanics and in classical particle mechanics. Anyhow, these are different theories. The sets of *global links* (GL) represent the intertheoretical connections between such theories.

One central methodological claim of the structuralist view is that, after a logical reconstruction of some theories under consideration, we gain results about their relations to other theories. In intertheoretical relations it is possible to identify structures that might appear in both related theories. The respective potential models (i.e., their general frameworks) of different theory-elements can be related through such relations.

As an example of a description of the logical structure of an empirical theory in the structuralist view, let us consider the potential model Mp of classical collision mechanics (see Balzer et al. 1987, 26–27):

Mp(MCC): x is a potential classical collision mechanics
(x ∈ Mp(CCM)) iff there exist: P, T, v, m, such that:

1. $x = \langle P, T, v, m \rangle$
2. P is finite and non-empty
3. T contains exactly two elements $(T = \{t_1, t_2\})$
4. v: $P \times T \rightarrow R^3$
5. m:$P \rightarrow R^+$

P is a set of discrete bodies that can be called particles. T is a set of time instants. v is the velocity function, assigning to each particle p and point of time its velocity as an element of R^3. Velocity is a time-dependent vectorial function whose ranges are triples of real numbers. It assigns a three-component vector (one component for each direction in space) to each particle at each time. m is the mass function, assigning to each particle its mass.

This, I argue, is methodologically close to Carnap's method of purely structural definite descriptions. It is important to keep in mind that when Carnap develop his method, no developed set theory or model-theory existed. And it is also known that, as outlined above, Carnap focused on structural descriptions of our knowledge of the world in a direct sense, i.e., by means of an ideal epistemic agent. The structuralist view starts by describing the structure of our empirical theories. These are two different starting points, but even so, both share a structuralist commitment to the description of our knowledge of the world. This is where I see a strong methodological analogy, both in the formal tools that are applied and in the motivation of providing objectivity through a focus on structural descriptions. In the structuralist view, such descriptions are never called structural descriptions but set-theoretic predicates, or *models*. Nevertheless, if we focus on the concrete tools that are applied, the connection between both should become clear.

Questions of scientific realism are scarcely addressed in the structuralist view. The adherents of this view aim to be neutral with respect to such questions. They aim to provide objective formal explications of the dynamics of empirical theories and of their logical structure. This is a point in which the structuralist view and Carnap are in total agreement. Though Carnap is quite explicit with respect to such questions, in the structuralist view there is no textual evidence of such a view.[2] Carnap expresses his neutralism at many passages in the *Aufbau*, such as here in §5:

[2]The supposed neutrality of the structuralist view has been mentioned to me several times in personal conversations by several of its leading adherents, such as C. Ulises Moulines, Wolfgang Balzer, Pablo Lorenzano, and José Díez.

> Are the constituted forms "generated by thought", as the Marburg-School teaches, or are they "only recognized", as realism states? Constitutional theory uses a neutral language; and there, forms are neither "generated", nor are they "recognized", but "constituted"; and it should be emphasized strongly already here that the word "constitute" is always meant completely neutral.

Kuhn and the Structuralist View

Let us now consider the connections between the structuralist view and Kuhn's (1962/1970) view on theory change. I first briefly mention the core ideas of Kuhn's proposal of theory change before we come to the outline of the connection to the structuralist view.

In Kuhn's system, a scientific community is a group of people that share and use the same paradigms. He introduced the notion of normal science as the activity of "puzzle-solving." In normal science, scientific research is guided by a paradigm. Anomalies can occur; this can lead to a crisis in certain fields. Such a crisis could—but must not—lead to what Kuhn called extraordinary science.

If a crisis leads to extraordinary science, it happens that one paradigm is substituted by a new one, and, after some time, a scientific revolution occurs. The scientists that were applying the old paradigm can no longer successfully communicate with the scientists applying the new paradigm.

Furthermore, Kuhn described the four components of a paradigm as follows:

1. *Symbolic generalizations*: In order to comprise knowledge, certain symbols are introduced and generalized. One concrete example are equations as they occur in practically all fields of science.
2. *Models*: There are heuristic models and ontological models. Heuristic models are mere fictions; the ontological models correspond partially with the world. For example, we imagine that planets and stars are actually round, but we do so only for practical reasons.
3. *Values*: The methodological values guide the scientific research and raise questions of technological applicability, ethic questions, and questions of the coherence of the research. For example, certain research areas might not be addressed for ethical reasons (genetic engineering, nano-technology, etc.). Another example is that, normally, scientists and philosophers of science accept the methodological values of theory-simplicity, formal elegance, coherence, and economy.
4. *Exemplars*: These are the paradigmatic applications, the concrete instantiations of a paradigm. Such concrete cases show how a paradigm actually works. These are the particularly efficient intended applications of a paradigm.

We focus now on the relation between Kuhn and the adherents of the structuralist view. We see that, despite Kuhn being the one who influenced the development of the structuralist view, he was also influenced by others. Sneed (1971) explicitly mentions the relation between his work and Kuhn's proposals as follows:

> It is certainly plausible to think that the initial successful application of the core of the theory is essentially the same for all those who have the theory. Different people who have the theory at a later time in its development may believe different statements. They may be more or less clever in seeing ways to extend the theory, and more or less successful in convincing their colleagues what evidence supports the claims they make with the theory… it is quite clear that Kuhn's thesis strongly suggests that we should modify our notion of what it is to have a theory of mathematical physics so as, at least, to require that everyone who has the theory has it "because of" the same initial success (ibid., 292–293).

We can see here that Sneed recognizes the fact of theory change and aims to include it in his framework. Sneed and the other developers of the structuralist view did exactly this; they modified the notion of what it means to have a theory of mathematical physics. As we have seen above, by identifying a theory with the formal notion of structure, a first step towards such a Kuhnian aim is taken. It is clear that one formal representation of a model of classical collision mechanics does not suffice to incorporate Kuhn's thesis of theory change into the structuralist framework, but within the structuralist view it was a first step towards an incorporation of Kuhn's ideas. This incorporation became stronger once the structuralist view developed. Today, there exist many formal notions that serve to model Kuhn's ideas. The most sophisticated work on this is Moulines (2011), where four types of theoretical development are achieved, with the mention of concrete historical examples. In the same passage, Sneed moves on and says:

> Again in Kuhn's terminology, we have said very little about "scientific revolutions" as they occur in mathematical physics. I confess, at the outset, that this is a subject about which I find it extremely difficult to say anything that is both precise and interesting. Nevertheless, the view of the logical structure of theories of mathematical physics I have been defending does appear to have some consequences relevant to such questions… (ibid., 296).

Sneed expresses his lack of interest in questions of radical theory change. Nevertheless, he is clear that he thinks it is his apparatus of set-theoretic formal reconstruction of physical theory that leads to a better understanding of Kuhn. Beside the formal aspect, in a wider philosophical sense, Stegmüller's (1976) notion of a relativized a priori reveals a close connection to Kuhn. Stegmüller expresses this as follows:

> The reason that we may only speak of a relative a priori is that no core, be it ever so sophisticated and yielding many successful expansions, can be guaranteed never to get caught in an a priori conflict with some future alternative and go down before it because this opponent can "deal with anomalies which it cannot"… Kant claimed that his theory reconciled rationalism and empiricism, the a priori and the empirical components in the scientific process. The reconstruction of Kuhnian theory dynamics with Sneed's conceptual apparatus is perhaps a better candidate for this job (ibid, 218).

The relativized a priori, expressed in the formal language of the structuralist view, is the theory-core K. However, what is actually subject to changes are the intended applications of a theory. Besides Stegmüller, Friedman (2001) also mentions a relativized a priori in the philosophy of science. His conclusion about a stable core, which can also undergo changes, is very similar to the view advocated by Stegmüller.

As mentioned above, Kuhn (1976) himself refers to the structuralist view and its influence on his view, and he recognizes explicitly the enriching contribution to his program provided by Sneed, and in a more systematic way by Stegmüller:

> To a far greater extent and also far more naturally than any previous mode of formalization, Sneed's lends itself to the reconstruction of theory dynamics, the process by which theories change and grow… Sneed also suggests and Stegmüller elaborates the possibility that at least some cases of change of core correspond to what I have called scientific revolutions… Though the Sneed formalism does permit the existence of revolutions, it currently does virtually nothing to clarify the nature of revolutionary change. I see, however, no reason why it cannot be made to do so, and I mean here to be making a contribution toward that end (ibid, 184).

Kuhn leaves it open for future work to develop the structuralist view more in the sense of his proposals. He recognizes explicitly that it is the structuralist view as a formal approach in the philosophy of science that pays tribute to his program and that allows his ideas to be expressed in a formal framework.

Conclusion

In the *Aufbau*, the main aim is to provide a logical method for the reconstruction of our knowledge of the world. Such a description is provided by purely structural (relational) descriptions. The structuralist view aims to describe the logical structure and the dynamics of scientific theories. As scientific theories are taken to be our most sophisticated, elaborate, and systematized descriptions of our knowledge of the world, it is, as in Carnap's, a proposal for reconstructing our knowledge of the world. It just starts from the reconstruction of our empirical theories and is, in this sense, not a direct but an indirect description. Kuhn's ideas about revolutionary theory change are reformulated in a logically precise sense in structuralism. Within the framework of structuralism, it is formally visible how theories actually change through time and how these are interrelated in every single case. The dynamics of scientific theories are modeled logically, not only metaphorically, as in Kuhn. Stegmüller alludes to a relative a priori, which he associates with the structuralist conception of what a scientific theory is. The whole structuralist program is addressing a wider range of questions than modeling theoretical change. However, one core part of Stegmüller's contribution to the development of structuralism is his analysis of Kuhn's ideas on theory change and also his application of a structural view of our knowledge about empirical theories. It is clear that at no point does structuralism refer explicitly to Carnap's *Aufbau* as their primary source of information and motivation. We have seen that the direct motivation for the methodological tools for structuralism can rather be found in Suppes' (1957) method of defining set-theoretic predicates. Notwithstanding, there are many indices to suppose an "indirect" connection between the early Carnap and the structuralist program. Although Carnap's logical method of expressing our knowledge in purely structural terms is strictly speaking not equivalent to what in structuralism is usually

taken to be the right logical tool for analyzing empirical theories, in both approaches there is a primary focus on structures. Carnap aims to describe our knowledge of the world in purely relational terms. Structuralism describes our knowledge of scientific theories in structural terms.

References

Balzer, W., Moulines, C. U., & Sneed, J. (1987). *An architectonic for science*. Dordrecht: Reidel.
Carnap, R. (1928/1998). *Der logische Aufbau der Welt*. Repr. Hamburg: Felix Meiner.
Friedman, M. (2001). *Dynamics of reason. The 1999 Kant Lectures at Stanford University.* Stanford: CSLI Press.
Kuhn, T. (1962/1970). *The structure of scientific revolutions.* 3rd ed. Chicago: University of Chicago Press.
Moulines, C. U. (1973). *La estructura del mundo sensible. Sistemas fenomenalistas*. Barcelona: Ediciones Ariel.
Moulines, C.U. (1991). Making sense of Carnap's Aufbau. *Erkenntnis*, *35*, 263–286.
Moulines, C.U. (2008). Die Entwicklung der modernen Wissenschaftstheorie. *Eine historische Einführung*. LIT-Verlag, Hamburg. 1890–2000.
Moulines, C. U. (2011). Cuatro tipos de desarrollo teórico en las ciencias empíricas. *Metatheoria, 1*(2), 11–27.
Richardson, A. (1998). *Carnap's construction of the world. The Aufbau and the emergence of logical empiricism*. Cambridge: Cambridge University Press.
Sneed, J. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.
Stegmüller, W. (1969). *Der Phänomenalismus und seine Schwierigkeiten—Sprache und Logik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II. Theorie und Erfahrung. Zweiter Halbband. Theorienstrukturen und Theoriendynamik* (Vol. II). Berlin: Springer.
Stegmüller, W. (1976). *The structure and dynamics of theories*. Berlin: Springer.
Suppes, P. (1957/1999). *Introduction to logic*. 2nd ed. New York: Dover.
Kuhn, T. S. (1976). Theory-Change as Structure-Change: Comments on the Sneed Formalism. *Erkenntnis*, *10*, 179–199.

Author Biography

Thomas Meier is a member of the Munich Center for Mathematical Philosophy (MCMP), at the LMU Munich. Thomas studied Philosophy and Linguistics in Mexico, Munich and Miami. He wrote his PhD thesis at the MCMP under the supervision of Carlos Ulises Moulines, Hannes Leitgeb, Dietmar Zaefferer and Otávio Bueno. He currently works on the research project "Frameworks for Rationality—The Debate on Scientific Realism and its Varieties" in collaboration with Otávio Bueno (Miami).

Part V
Various Echoes

Abstraction and Epistemic Economy

Marco Panza

Introduction

Most[1] arguments usually supporting the view that some abstraction principles are analytic depend on ascribing to them some sort of existential parsimony or ontological neutrality,[2] whereas the opposite arguments, aiming to deny this view, contend this ascription. As a result, other virtues that these principles might have are often overlooked. Among them, there is an epistemic virtue which I take these principles to have, when regarded in the appropriate settings, and which I suggest be called "epistemic economy". My present purpose is to isolate and clarify this notion. I also try to make clear that complying with this virtue is essentially independent of complying with existential parsimony or ontological neutrality.

The intimate connection between the analyticity of an abstraction principle and its existential parsimony or ontological neutrality can be questioned and, in my view, the analyticity of such a principle can be made to depend on its epistemic economy instead. Hence, distinguishing epistemic economy from existential parsimony and/or ontological neutrality would allow a denial that an abstraction principle is existentially parsimonious or ontologically neutral, or keeps an agnostic view on the matter, to maintain nevertheless that it is analytic, according to some plausible construal of analyticity.

[1]I thank Andrew Arana, Jeremy Avigad, Francesca Boccuni, Annalisa Coliva, Sorin Costreie, Michael Detlefsen, Sébastien Gandon, Guido Gherardi, Emmylou Haffner, Bob Hale, Brice Halimi, Greeg Landini, Paolo Mancosu, Ken Manders, Daniele Molinini, Julien Ross, Andrea Sereni, Stewart Shapiro, Giorgio Venturi, Sean Walsh, and David Waszek for useful comments and/or suggestions.

[2]I am not interested here in discussing the relation between existential parsimony and ontological neutrality. All I shall say of these virtues is independent of this matter.

M. Panza (✉)
CNRS, IHPST, University of Paris 1, Panthéon-Sorbonne, Paris, France
e-mail: panzam10@gmail.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_17

In my view, an abstraction principle (similar to any other axiom or definition) is epistemically economic not because of its logical nature, nor because of some of its other intrinsic features, nor even because of its being immersed in certain logical systems, but rather because of its context of use, that is, the setting and purpose in and for which it is used, when immersed in such logical systems.

In particular, I limit my attention to the use of abstraction principles in formal definitional contexts, namely to their being involved in different formal definitions of natural and real numbers (and, incidentally, of integer and rational ones). These definitions are all complex in the sense that they do not depend merely on a single stipulation but rather on a system of stipulations of different sorts. The abstraction principles I take into account are among such stipulations, and I take their being epistemically economic (or not) as being the same as their being involved in definitions which are themselves epistemically economic (or not). Insofar as all these complex definitions also include explicit definitions, I then regard the relevant abstraction principles being or not being epistemically economic on par with these explicit definitions being or not being so. Broadly speaking, I take a definition, and the abstraction principles possibly involved in it, to be epistemically economic if its understanding involves less and/or more basic intellectual resources than other relevant definitions of the same items, or, in case the required resources are (nearly) the same, if its understanding is more progressive than that of other relevant definitions of the same items.[3]

This explanation is quite rough. In the next section, I try to elucidate the notion of epistemic economy of a formal definition in general. This is a hard task, however. To achieve it, one should explain, in general, intricate notions such as those of understanding and of intellectual resources to be deployed to achieve understanding, as well as provide a way to compare the amounts of the intellectual resources involved in understanding different definitions of the same items, and their being more or less basic. It is therefore not surprising that I won't be able to get a general, clear-cut, unequivocal and compact characterisation of this virtue. I merely hope to make my general idea reasonably clear, so as to open the road for further enquiries.

In the following sections I look at the matter more *in concreto*, as it were, by considering different examples and dealing with them in comparative terms. I will firstly briefly take stock of (a part of) the discussion on the analyticity of Hume's principle (HP), by mainly discussing a neo-logicist argument according to which HP contrasts with Peano axioms insofar as the latter provide an arrogant implicit definition of natural numbers, whereas the definition provided by the former avoids arrogance. This is intended to clarify that and how this discussion essentially focuses on existential issues. In reaction to this, I will then suggest another way to contrast the same definitions, by comparatively assessing their epistemic costs, and

[3]Plausibly, epistemic economy not only applies to definitions or parts of them. However, apart from some short remarks on the epistemic cost of proofs in the next section, I do not investigate this matter further here.

I will argue that Frege Arithmetic (FA)—the neo-logicist version of second-order arithmetic, involving HP as an axiom—supplies an epistemically economic definition of natural numbers. Later, I will compare four definitions of real numbers, three of which involve abstraction principles, and argue that one of the latter is epistemically economic. Finally, I will offer some concluding remarks.

Epistemic Economy

Speaking of epistemic economy might bring to mind Mach's idea of economy of though (Mach 1883, § V.4, pp. 452–466, esp. § V.4.6, pp. 460–461). E.C. Banks has distinguished two doctrines concerned with this idea (Banks 2004, p. 24). The first relates to the way in which "science structures its laws under one another to maximise desirable features", by grouping "the greatest number of particular experiences under the least number of super categories and principles". This doctrine is "descriptive", because, for Mach, "nature […] [is] lawlike", i.e., "objective temporal and spatial patterns exist […] in nature ready to be arranged under one another". The second doctrine is normative, instead, because it pertains to "the role of economy in the framing of basic laws". However, normativity here essentially depends on our cognitive faculties, mainly memory, because "the emergence of general concepts and laws" is explained by Mach in terms of "memory's operation over traces"(ibid., p. 31). This second doctrine was the target of Husserl's allegation of psychologism (Husserl 1900, Chap. 9). According to Husserl, when speaking of economy of thought, Mach was "ultimately" bearing on "a branch of the theory of evolution", with the result that his "attempts to found epistemology on an economy of thought ultimately reduce[d] to attempts to found it on psychology" (Husserl 2001, vol. I., p. 128). To these attempts, Husserl opposed a different program. He focussed on the "thought-economy which occurs in the purely mathematical discipline, when genuine thought is replaced by surrogative, signitive thinking", from which "almost without specially directed mental labour, deductive disciplines arise having an infinitely enlarged horizon", and he envisioned undertaking a detailed investigation of the different methods allowing this "economic achievement"(ibid., pp. 127–128).

What I mean by epistemic economy stands between Mach's descriptive doctrine and Husserl's program. On the one hand, I do not endorse the claim that objective patterns (whatever they might be) already exist in nature "ready to be arranged under one another", but I agree that a scientific theory, as well as a mathematical one, results from a certain way of structuring appropriate prior material. On the other hand, I do not endorse the ideas that, in pure mathematics, "genuine thought is replaced by surrogative signitive thinking", and that "deductive disciplines" require almost no "specially directed mental labour", but I agree that formal mathematical theories use signs, or, better, appropriate formal languages, to render a previous informal thinking.

To be more precise, I take mathematics to result from our intellectual activity and consequently consider that achieving a mathematical task depends on deploying some intellectual resources. I view formal theories as convenient means for expressing and controlling abstract thinking. More precisely, I consider that the purpose of a formal theory is to re-cast a certain piece of our informal knowledge in such a way that the different ingredients involved in this theory are so arranged that it becomes transparent what rests on what. Hence, understanding a formal theory, or any component of it (for example a definition, a theorem, or a proof), consists, in my view, of recognising it as re-casting the relevant pieces of some informal knowledge in such a way that its different ingredients—which are, in turn, to be recognised as re-casting some pieces of our informal knowledge—obey a certain arrangement. Moreover, I take the intellectual resources involved in this understanding to be those that are required to achieve this recognition. Thus, a formal theory or any of its components are, in my view, epistemically economic if they are so shaped that achieving such a recognition in the relevant case calls for less and/or more basic intellectual resources than in the case of other theories or comparable components of them, or, if these resources are (nearly) the same in both cases, when they come about in the former more progressively than in the latter.[4] This means that the former theory (or component) re-casts the relevant piece of our informal knowledge by also re-casting for this purpose less and/or more basic other pieces of our informal knowledge than the latter theory, or do it by re-casting more progressively (nearly) the same pieces of our informal knowledge.

One could go further and maintain that a formal theory, or any of its components are as epistemically economic as possible, or utterly epistemically economic if achieving such a recognition relatively to them calls for as few and/or as basic intellectual resources as possible, that is, if they re-cast the relevant piece of our informal knowledge by also re-casting for this purpose as few and/or as basic other pieces of our informal knowledge as possible. Determining whether this last condition occurs would require a modal appraisal that could be difficult. I therefore conform to the former, weaker, account.

As an example (to which I will come back later), consider FA. It is intended to re-cast informal arithmetic, and it appeals, for this purpose, to an appropriate re-casting of several informal notions, such as those of an object and of a first-level concept,[5] that of identity for objects, that of the falling of an object under a

[4]By this, I mean that the relevant resources, or a significant part of them, come about in the former case in agreement with an order in which some conceptually depend on others but not vice versa, whereas they come about in the latter case all together at once.

[5]Here, I adopt the current neo-logicist interpretation of monadic predicate variables as ranging over first-level concepts. *Mutatis mutandis*, one could go for another option and interpret such variables as ranging over properties of objects. If first-level concepts are conceived as they are by neo-logicists, then it does not seem to me that this would make any significant difference.

(first-level) concept, etc. (I will offer later a more comprehensive inventory). For short, we could say—and, indeed, I use this way of speaking, in what follows—that understanding FA calls for these notions. This means that understanding FA depends on recognising it as re-casting informal arithmetic by also re-casting all these notions according to a certain arrangement. Arguing that FA is epistemically economic is the same as arguing that these notions are fewer and/or more basic than those on the re-casting of which the understanding of other current formal theories (especially second-order ones), which are also intended to re-cast informal arithmetic, depends on. This means that FA achieves this task by also re-casting for this purpose fewer and/or more basic informal notions than these other theories.

It appears to me that, thus conceived, the notion of epistemic economy comes close, although in a different setting, to one of Frege's crucial concerns.

Famously, Frege considered a truth to be analytic if, in its proof, "one only runs into logical laws and definitions" (Frege 1884, § 3; I slightly modify the translation offered in Frege 1953). By 'definitions' Frege clearly meant explicit definitions, and, for him, an explicit definition could certainly not be used to introduce new items and, a fortiori, for positing their existence. Together with the presently common view that a logical law can have no existential import, this could lead to the conclusion that Frege's pursuit of analyticity is a pursuit of existential parsimony. However, his conception of logic[6] suggests another view. He distinguished second-order logic as such, neither from first-order, nor from propositional logic, thus failing to grasp (or avoiding to emphasise) the essential difference (mainly in matters of ontology) we see between them. Moreover, he did not conceive of comprehension as a specific condition to be met through the admission of specific axioms, but rather came to results analogous to those we obtain by comprehension through the apparently innocent adoption of unrestricted rules of substitution in language. Finally, he candidly embraced the notion of a logical object. He took his logical language to be meaningful, and included under his (quite generic) notion of a law of logic both statements written in what is for us a purely logical language, and others involving what is for us a non-logical vocabulary. He was then perfectly comfortable with the idea that fixing the laws of logic comes together with ensuring (or revealing) the existence of some logical objects. Hence, restricting a proof to rely only on logical laws and (explicit) definitions was, for Frege, less a way of pursuing existential parsimony than of limiting the tools to be used in conducting a proof. A similar concern also appears both in his notion of a priori—according to which a truth is a priori if it admits a proof that "proceeds as a whole from general laws which neither need nor admit a proof"—and in his claim that the question of whether a truth is a priori or a posteriori, analytic or synthetic is settled by "finding […][its] proof and tracking it back up to the original truth" (ibid.). Frege's main

[6]Regarding my appraisal of Frege's conception of logic, I refer the reader to Cadet and Panza (2015).

purpose was, then, that of tracking arithmetical truths back to the minimal tools required to prove them.

This was for him a way to identify the place of these truths within the objective general order of truths, which he was aiming to reconstruct. Once the idea that there exists such an objective order is dismissed, and epistemology is no longer understood to be concerned with the reconstruction of such a putative order, but rather with the activity that human subjects perform in order to constitute a body of knowledge, and with the resources to be deployed to this end, probative tools appear as part of these resources, and minimising the former results in an effort to economise the latter. Moreover, once epistemology is so conceived, proofs appear less as procedures for discovering actual truths than as arguments for obtaining conclusions starting from some assumptions or from other previously established conclusions. Hence, assessing the tools required to conduct a proof appears less as a way to estimate the objective role of a certain truth than as a way to evaluate the epistemic cost of reaching a certain conclusion within the relevant theory, namely the intellectual resources that are to be deployed within this theory to fix a certain thought. When replaced in such a framework, his concern for whether a truth is a priori or a posteriori, analytic or synthetic, comes close to a concern for epistemic economy.

Still, whereas Frege's concern seems to point to the epistemic cost of the proofs involved in the relevant theories, that is, to the resources to be deployed to understand these proofs, I rather focus on the epistemic cost of the definitions that make these proofs possible, that is, on the resources to be deployed to understand these definitions, independently of the proofs they give rise to.

Arrogant Versus Non-arrogant Implicit Definitions

I begin my enquiry by considering the neo-logicist definition of natural numbers.

Even though most discussions of this definition focus on HP, there is more in the former than the mere stipulation of the latter. For the definition to work, a previous definition of objects and first-level concepts is required. Immersing HP in a suitable system of second-order logic simply makes this definition available. Whether explicitly mentioned or not, objects and first-level concepts are taken in the former definition to be implicitly defined by this system, as the items which its individual and monadic predicate variables are supposed to range over, respectively, by admitting that the existence of the latter is ensured by comprehension. Once objects and first-level concepts are thus defined, HP acts as the implicit definition of a function taking the latter and giving a particular kind of the former, namely numbers of first-level concepts, that is, cardinal numbers. To go from these numbers to natural ones, one has to extend comprehension to formulas including the functional constant designating this function, and to rely on this extension to state some

explicit definitions, which allows one to single out the natural numbers from the cardinal ones.[7]

All this will be made clearer in the next section. For the time being, it is enough to remark that in this context, the neo-logicist thesis that HP is analytic admits two different readings. According to one of them, that which is claimed to be analytic is HP as such, merely conceived as providing an implicit definition of a function inputting first-level concepts and outputting cardinal numbers. According to the other reading, that which is claimed to be analytic is the whole system of stipulations providing the definition of natural numbers, i.e. the whole FA, including the explicit definitions allowing natural numbers to be singled out among cardinal ones. For brevity, let us term this thesis 'weak neo-logicist analyticity thesis' if it is taken under the former reading, and 'strong neo-logicist analyticity thesis', if it is taken under the latter one. Insofar as FA involves comprehension extended to formulas including the functional constant introduced by HP, the weak thesis appears, at least at first glance, more plausible than the strong one. Still, it is only by endorsing the strong one that it can be argued that the neo-logicist definition of natural numbers results in a vindication of Frege's logicist program, as neo-logicists claim. It seems therefore, that whatever the arguments for or against the analyticity of HP might be, for them to be fully relevant for our appreciation of this program, and, more generally, for the philosophy of arithmetic, they have to be for or against the strong thesis.

This is the case of the argument against the analyticity of HP which is most often repeated and considered as convincing. In short, its point is that HP cannot be analytic because for it to hold there must be infinitely many objects (Boolos 1997, p. 306).[8] However, as remarked by Boolos himself, there is something biased in this argument, because, in Boolos's words, "one person's *tollens* is another's *ponens*, and Wright happily regards the existence of infinitely many objects, and indeed, that of a Dedekind infinite concept, as analytic, because they are logical consequences of what he takes to be an analytic truth" (ibid.). This is reminiscent of Frege's view that natural numbers are logical objects because their existence is required for some proper names to refer, and these names actually refer because this is in turn required for some logical truths to be true.

However, one might retort, if natural numbers are logical objects or their existence analytic, then (second-order) Peano axioms, when conceived as categorical

[7]Taking the relevant system of second-order logic to implicitly define first-level concepts (or properties of objects: cf Footnote 5) is not mandatory. One might merely take HP as an implicit definition of a functional constant inputting monadic predicates and outputting terms. However, in this case, adding this principle to the axioms of such a system of logic would merely result in introducing appropriate numerals, rather than in defining cardinal numbers. These numerals being given, defining natural numbers would then require much more than merely singling the latter out among the former, because this selection would at most provide a family of terms.

[8]This argument questions the strong neo-logicist thesis because it is only in the presence of second-order logic with comprehension appropriately extended, and of appropriate explicit definitions, that the existence of infinitely many objects follows from HP (through the admission of appropriate explicit definitions).

implicit definitions of these numbers, should be logical or analytic truths. And if Peano axioms are so, why should appealing to HP to define natural numbers be preferable to merely stating these axioms? The obvious answer is that defining natural numbers though HP allows us to recognise that natural numbers are logical objects or their existence analytic, as it is the possibility of this very definition that makes it so, whereas this is not the case for the definition of natural numbers appealing to Peano axioms.[9]

This could lead one to think that the crucial point under discussion does not concern existential issues but rather an alleged intrinsic difference between HP and Peano axioms—namely, the former having a virtue (apparently an epistemic one) that the latter do not have—which does not depend on their respective truth requiring or not requiring the existence of some objects, that is, on their having or not having an ontological import for objects.[10] However, when one looks at the situation more carefully, this appearance dissipates.

To see this, consider how Hale and Wright account for the relevant difference they see between HP and Peano axioms. Their main point is that the definition of natural numbers provided by the latter is arrogant, whereas that provided by the former is not. The point is made on different occasions, but it receives particular emphasis in Hale and Wright (2000, 2009).

The former paper mainly focuses on implicit definitions. These are taken to include two parts: an unsaturated "matrix" formed by "previously understood vocabulary" (or possibly only by logical-constants and variables or schematic letters), and the *defininedum* or *definienda* to be inserted in this matrix, so as to form a well formed sentence or system of sentences (ibid. pp. 285, 289).[11] It is then arrogant if "the antecedent meaning" of the matrix and "the syntactic type" of the *definiendum* or *definienda* are such that the truth of the relevant sentence or sentences "cannot justifiably be affirmed without a collateral (a posteriori) epistemic work" (ibid., p. 297).

To this general characterisation, a sufficient and necessary condition for an implicit definition not to be arrogant is added. If I understand it correctly, the former (generally stated at pp. 314–315) goes as follows: let 'f' designate a *definiendum* and let '$\boldsymbol{S}(f)$', '$\boldsymbol{S}_I$' and '$\boldsymbol{S}_E$' be three appropriate sentences or schemes of sentences, the first of which includes one or more occurrences of 'f', whereas the other two include no occurrence of 'f' and of any constant designating another *definiendum*;

[9]Of course, the existence of natural numbers being analytic is strictly not the same as their being logical objects, as well as a truth being logical is not the same as its being analytic. These important distinctions are not relevant for the issue under discussion so it is not necessary to insist on them here.

[10]In short, I say that a stipulation, or a system of stipulations, has an ontological import for objects if the truth of this stipulation, or these stipulations, requires the existence of some objects.

[11]As a matter of fact, Hale and Wright only consider the case of definitions given by a single sentence including a single *definiendum*, but the generalisation to the case of systems (or conjunctions) of sentences including more *definienda* is as natural as it is necessary to adapt the account to the case of Peano axioms.

and the third does not "introduce [any] fresh commitments"; then, for an implicit definition of f not to be arrogant, it suffices that it results from stipulating the truth either of '$\boldsymbol{S}_I \Rightarrow \boldsymbol{S}(f)$', or of '$\boldsymbol{S}(f) \Rightarrow \boldsymbol{S}_E$', or of both (ibid., pp. 299, 302). Hale and Wright seem, here, to imply that the mere conditional form of these sentences or schemes of sentences is enough to ensure that their truth can "justifiably be affirmed without a collateral (a posteriori) epistemic work". Why this is so? The necessary condition suggests a response, at least for the case to which it applies. According to it, to avoid arrogance "the stipulation of the relevant sentence as true ought not to require reference for any of its ingredient terms in any way that cannot be ensured just by their possessing a sense" (ibid., p. 314). It seems, then, that, according to Hale and Wright, in the cases where '$\boldsymbol{S}(f)$' includes some constant terms involving 'f', possibly schematic ones (that is, 'f' is either an individual or a functional constant), the conditional form of '$\boldsymbol{S}_I \Rightarrow \boldsymbol{S}(f)$' and '$\boldsymbol{S}(f) \Rightarrow \boldsymbol{S}_E$' ensures that the truth of these very sentences or of their relevant instances does not require (as a necessary condition) that these terms or their corresponding instances refer in a way not possibly ensured by their acquiring a sense thanks to the very stipulation of such a truth. This means that it is not a necessary condition for these sentences or their relevant instances to be true that some objects exist. Hence, the relevant "collateral (a posteriori) epistemic work" is that which would be required to ensure that these objects exist.

This point well applies to HP and Peano axioms (perhaps too well to not generate the suspicion that it is taken ad hoc). Suppose that 'f' stands for the functional constant '$\#$', '$\boldsymbol{S}(-)$' stands for '$-P = -Q$', and '$\boldsymbol{S}_I$' and '$\boldsymbol{S}_E$' both stand for '$P \approx Q$' (where 'P' and 'Q' are schematic monadic predicates). From the sufficient condition, it follows that stipulating the truth of HP—namely, '$\#P = \#Q \Leftrightarrow P \approx Q$' (where '$P \approx Q$' abbreviates a formula of second-order logic asserting that the objects falling under P and those falling under Q are in bijection)—provides a non-arrogant implicit definition of the function $\#$, because, for '$\#P = \#Q \Leftrightarrow P \approx Q$' to be true, it is not necessary that '$\#P = \#Q$' be true, in turn, and, then, that the relevant instances of '$\#P$' and '$\#Q$' refer in the required way. On the other hand, for any Peano axiom—for example '$Suc(n) \neq 0$' (where 'n' is a schematic individual constant)—to be true, it is necessary that the constant terms included in it, or their relevant instances—namely '0' and the relevant instances of '$Suc(n)$'—refer in this way. From the necessary condition, it follows, then, that stipulating the truth of Peano axioms results in an arrogant implicit definition.

An easy reply to this argument has been suggested by MacFarlane (2009, pp. 454–455). Let PA be the conjunction of all Peano axioms (in some appropriate form), and CPA the double implication '$\text{PA} \Leftrightarrow \forall x(x = x)$'. Clearly, CPA is as conditional as HP is, but stipulating its truth has the very same consequences as stipulating PA's truth. One could retort that CPA "makes the existence of [natural] numbers conditional on a logical truth" and is then conditional in a "Pickwickian sense". Certainly. However, "HP, too, makes the existence of [natural] numbers conditional on logical truths [, and][…] that is precisely why it can serve as the basis of a kind of logicism". Moreover, both the left-hand sides of CPA and of HP

have "ontological commitments "of which their right-hand sides are "innocent". Hence, if defining natural numbers through HP avoids arrogance because of the conditional form of this principle, then defining natural numbers through CPA should also avoid arrogance.

The answer might be too easy, because the right-hand side of HP is not, as such, a logical truth; only some of its relevant instances are so. Hence, the truth-conditions of HP do not reduce to the truth-conditions of its left-hand side, as is the case for CPA. This is precisely what Hale and Wright retort to MacFarlane's objection in Hale and Wright (2009): any instance of HP whose right-hand side is a logical truth "is to be viewed as part of a package of stipulations whose role is to fix the truth-conditions of statements of numerical identity"; what matters, then, are these conditions, namely the fact that they are "as feasible as any other purely meaning-conferring stipulations", which entails that "there is no need—indeed no room—for any associated stipulation of numerical existence" (Hale and Wright 2009, p. 476).

Here, as in the argument of Hale and Wright (2000), the essential point is that HP has, a such, no ontological import for objects, because its truth does not require that the relevant instances of '$\#P$' and '$\#Q$' refer in a way that is not ensured by their acquiring a sense, thanks to the stipulation of this truth, and, then, that cardinal numbers exist, whereas Peano axioms, as well as their conditionalisation CPA have such an import, because their truth requires that '0' and the relevant instance of '$Suc(n)$' refer in this way, and then that natural numbers exist. To complete the argument, it is then enough to add that, once the truth of HP is stipulated, the existence of cardinal numbers is revealed by the revelation of the truth of appropriate instances of '$P \approx Q$', and that this truth is nothing but a logical truth (the existence of 0 is, for example, revealed by defining it as the cardinal number $\#[x: x \neq x]$ and by deriving the truth of '$0 = 0$' from HP and '$[x: x \neq x] \approx [x: x \neq x]$', which is a logical truth).

MacFarlane's objection does not come alone, however. It is part of a more general argument intended to show that PA fares as well as HP in all the requirements, other than non-arrogance, that Hale and Wright advance in Hale (2000) for an implicit definition to be acceptable as a meaning-conferring stipulation, namely consistency, conservativeness, generality, and harmony—supposing that the non-arrogance constraint is independent of these, which MacFarlane questions, at least insofar as non-arrogance reduces to conditionality, and Hale and Wright claim, instead: cf. MacFarlane (2009, p. 455) and Hale and Wright (2009, pp. 467–468).[12] In their reply, Hale and Wright question whether this is the case for

[12]As a matter of fact, MacFarlane's paper is also concerned with Hale and Wright's conception of numerical definite descriptions as singular terms in relation with the sort of logic that FA actually requires (namely whether this logic is classical or free). In Hale and Wright (2009), Hale and Wright also reply to this point, but, though somehow connected with the question I'm discussing, this matter can be left aside here.

the two last constraints (Hale and Wright 2009, pp. 466–467), but they do not insist on this point much. In Hale and Wright (2009), they mainly insist on the non-arrogance requirement to underline the difference between Peano axioms and HP, as alleged implicit definitions of natural numbers. They nonetheless seize the opportunity provided by this new paper to spread more light on this requirement.

They begin by suggesting both a new general characterisation of arrogance and a new sufficient condition for it. Their characterisation is the following: arrogance is "the situation where the truth of the vehicle of the stipulation is hostage to the obtaining of conditions of which it's reasonable to demand an independent assurance, so that the stipulation cannot justifiably be made in a spirit of confidence, 'for free' "(Hale and Wright 2009, p. 465). The sufficient condition is this: "a stipulation is arrogant just if there are extant considerations to mandate doubt, or agnosticism, about whether we are *capable* of bringing about truth merely by stipulation in the relevant case" (ibid., p. 468). By conversion, this implies that non-arrogant implicit definitions "are ones where there is no condition to which we commit ourselves in taking the vehicle to be true which we are not justified—either entitled or in possession of sufficient evidence—to take to obtain" (ibid.). The point is then concerned with the nature of the "independent assurance" that arrogant and non-arrogant explicit definitions do and do not require, respectively, and/or the reasons that can "mandate doubt, or agnosticism" about our capability of "bringing about truth", merely by making the relevant stipulations.

A new argument advanced by Hale and Wright might suggest that this nature not only pertains to existential considerations. Leaving their motivations apart, the claim is the following: "The stipulation of Hume['s principle] serves to communicate a singular-thought-enabling conception of the sort of objects the natural numbers are and explains their essential connection with the measure of cardinality. The stipulation of […] Peano [axioms] communicates no such conception, and actually adds no real conceptual information to what would be conveyed by a stipulation of their collective Ramsey sentence" (ibid., pp. 471–472). This is highly questionable, however, for reasons that do not directly depend on the reading of HP and of the consequent definition of natural numbers. What Hale and Wright base this conclusion on is their view that Peano axioms "convey no more than the collective structure of the finite cardinals—something which, because it entails those axioms, Hume['s principle] also implicitly conveys"; a view that they specify by claiming that these axioms convey "no conception of the sort of thing that zero and its suite are" (ibid., p. 471). Still, consenting to this requires admitting that there is something that natural numbers actually are, besides their forming a progression. In other words, what Hale and Wright take here for granted is not that there is room for coding these numbers with appropriate items having a determinate particular essence, but rather that these numbers have a determinate particular essence. It is only after having consented to this that the discussion can begin on whether HP conveys this essence, whereas Peano axioms do not. However, consenting to this is all but anodyne, and no argument for the definitional advantage of HP on Peano axioms can require this without being biased from the very beginning.

This might be the reason for Hale and Wright's rapid shift back to existential considerations. They remark that there is no concern over a possible replacement of HP with its Ramsification (whatever it might be), because "there is no need to (attempt to) stipulate that a suitable function [i.e. a function satisfying HP] exists", provided that "the existence of such a function is […] a consequence of something known as an effect of the stipulation, viz. Hume's Principle itself" (ibid, p. 473). The point here is that HP merely fixes "the truth-conditions of the canonical statements of numerical identity in which [the operator '#'][…] occurs" (ibid.). In other words, what really matters is that HP, as well as any abstraction principle suitable for working as an implicit definition, is "tantamount to legitimate schematic stipulations of truth conditions […][, whereas] to lay down […] Peano [axioms] as true is to stipulate, not truth-*conditions*, but *truth* itself" (ibid., p. 474). Hence, Hale and Wright go ahead, "as a stipulation that Hume['s principle] is considerably more modest than […] Peano [axioms]: the attempted stipulation of the truth of […] Peano [axioms] is effectively a stipulation of countable infinity, whereas whether or not Hume['s principle] carries that consequence is a function of the character of the logic in which it is embedded" (ibid., p. 475).

Even though it is enriched by a number of collateral considerations, the crucial argument seems to be the same as that already advanced in Hale and Wright (2000): what matters, concerning the definitional advantages of HP over Peano axioms, is that the latter have an ontological import for objects, better they entail the existence of an infinity of objects, whereas the former has no such import (and it works perfectly, as it is required to work in FA, without requiring that the existence of any function be admitted).[13]

For short, call this argument 'the existential argument'. Taken as such, it seems at most able to support the weak neo-logicist analyticity thesis, because it merely concerns HP, or better yet, nothing but the logical form of HP. Moreover, it is just part of the argument that HP acquires an ontological import for objects when it is embedded in an appropriate logical setting (and, I add, it is coupled with appropriate explicit definitions). Hence, to support the strong thesis, one should either advance an independent argument, or admit that the existential argument secures the weak thesis and infer the strong thesis from it. For this last purpose, one could admit: (i) that a system of stipulations is analytic, even though it has an ontological import for objects, if the objects whose existence is required by the truth of these stipulations (taken together) are such that their existence is to be regarded as analytic, in turn; (ii) that the existence of some objects is to be regarded as analytic if it follows from logic, plus some analytic principles, an appropriate extension of

[13]Neo-logicists have come back to this last point in different ways and in the context of different forms of argumentation. A very compact way to make the same point is found, for example, in Wright (1999), § II.2.

comprehension, and explicit definitions.[14] It would then be easy to conclude that the whole FA is analytic, though having an ontological import for objects, because HP, taken as such, is analytic. However, both these suppositions can be questioned. For example, one could argue that, for (ii) to be admissible, it should involve some conditions to be met by the relevant system of logic, so as to avoid that one be licensed to take as analytic a complex definition involving a system of logic whose innocence might be questioned. Once this is admitted, one could argue, in agreement with Shapiro and Weir (2000, §§ II and III), that the system of second-order logic involved in FA does not meet these conditions. This would block the derivation of the strong neo-logicist thesis from the weak one, with the result that the existential argument could not be taken as part of a larger argument in favour of the former thesis, even if it were regarded as suitable for supporting the latter.

This is not all. The suitability of the existential argument for supporting the weak thesis can be questioned, too. One could contend, for example, that from the fact that the truth of HP does not require the truth of '$\#P = \#Q$'—nor the existence of references for the relevant instances of '$\#P$' and '$\#Q$', i.e. of cardinal numbers—it does not follows that HP has no ontological import for objects. To this end, one could argue as follows. Insofar as the truth of the mere axioms of the system of second-order logic to which this principle is added to get FA does not require the existence of any object, the following options remain open: (a) this import is distributed among this system of second-order logic, HP taken as such, the extension of comprehension to formulas including '#', and the explicit definitions used to single out natural numbers among cardinals; (b) in the presence of this extension and of these explicit definitions, this system of second-order logic triggers the ontological import for objects of HP, without having itself any such import; (c) vice versa, in the presence of this extension and of these explicit definitions, HP triggers the ontological import for objects of this system of second-order logic, without having any such import itself; (d) taken together, both this system of second-order logic and HP trigger the ontological import for objects of this extension and of these explicit definitions, without having any such import themselves. The assessment of the ontological import of HP could provide a sound argument in favour of the weak neo-logicist analyticity thesis only insofar as it was suitable for positively supporting (c) or (d), or at least for discarding (a) and (b).

[14]Apart for the mention of the extension of comprehension (which neo-logicists seems to consider as existentially anodyne), condition (2) is suggested by Wright's following remarks (1999, pp. 307 and 310):

> Analyticity, whatever exactly it is, is presumably transmissible across logical consequence. So if second-order consequence is indeed a species of logical consequence, the analyticity of Hume's Principle would ensure the analyticity of arithmetic.
>
> […] on the classical account of analyticity the analytical truths are those which follow from logic and definitions. So if the existence of zero, one, etc. follows from logic plus Hume's Principle, then provided the latter has a status relevantly similar to that of a definition, it will be analytic, on the classical account, that *n* exists, for each finite cardinal *n*.

However, two problems arise. First, granting (c) or (d) would result in ascribing an ontological import for objects to the system of second-order logic, and to the extension of comprehension and the explicit definitions, respectively, and this would be at odds with the purpose of deriving the strong thesis from the weak one along the lines suggested above. Second, the existential argument is able neither to support positively (c) or (d), nor to discard (a) and (b), because all these options are left open by acknowledging that HP is not arrogant (insofar as its truth does not require the truth of '$\#P = \#Q$'), although contrasting the arrogance of Peano axioms with HP's avoiding arrogance is simply not relevant for choosing between the four options.

Moreover, one could also observe that deriving the existence of natural numbers in the neo-logicist setting hinges on the logical truth of appropriate instances of '$P \approx Q$' only in free logic. Because, if logic is not free, the mere introduction of the functional constant '#', via HP, and the extension of comprehension to formulas including this constant allow one to derive any instance of '$\#P = \#P$' from '$\forall x(x = x)$' (MacFarlane 2009, p. 447; that the logic underlying FA is to be free is also argued, among others, in Shapiro and Weir 2000, p. 108, and admitted by Hale and Wright in 2009, pp. 463–464). However, assuming that the relevant logic is free seems to be at odds with alleging that the existence of some objects is analytic if it follows from logic plus some analytic principles, a principle being analytic insofar as it has no ontological import for objects. Indeed, the adoption of free logic seems to be naturally linked with the view that existence cannot be a matter of logic.

I'm far from considering these or similar remarks as knock-down objections against the suitability of the existential argument for supporting the neo-logicist claim that HP is analytic, this claim being intended either in agreement with the weak or with the strong neo-logicist analyticity thesis. For my present purpose, it is enough to show that this argument can be questioned. This should be enough to urge anyone regarding the neo-logicist definition of natural numbers with interest to look for other reasons to favour it over other definitions, and, possibly, also to view HP, or better the whole FA, as analytic, namely reasons that are not based on the assessment of HP's ontological import for objects. My suggestion is that such a reason can be found in the epistemic economy of the former definition, which I regard as being independent of HP's having or not having this import.

Before arguing in favour of this suggestion a disclaimer is in order. The point I want to make is neither that the neo-logicist definition of natural numbers is definitely better than, or to be preferred to, any other current one, because it is epistemically economic nor that its being so makes this definition better than, or preferable to, any other current one from an epistemic point of view. My point is rather that its being so gives this definition a particular epistemic virtue that other current definitions do not possess, which makes the former preferable over the latter in appropriate circumstances and according to some aims—though I admit that other virtues, either epistemic or not, could be differently distributed among the relevant definitions, and could bring on other choices under different circumstances and according to different goals.

Comparing FA and Z_2 with Respect to Epistemic Economy

Until now, I followed Hale and Wright and spoke of Peano axioms in general (in fact they often speak of Dedekind-Peano axioms, but this makes no relevant difference). To settle some ideas, it is better to be more precise. In the present section, I consider Peano arithmetic under the form of Hilbert-Bernays second-order theory Z_2, in Simpson's version (Hilbert and Bernays 1934, suppl. IV; Simpson 2009), and compare its epistemic cost with FA's.

Z_2's language—let us call it '$\mathfrak{L}^{Z_2}$'—is two-sorted, and its two sorts of variables, 'i', 'j', 'k', 'm', 'n',… and 'X', 'Y', 'Z',… are intended to range over natural numbers and sets of natural numbers, respectively. $\mathfrak{L}^{Z_2}$ also includes the identity symbol for terms and six non-logical constants '0', '1', '$+$', '$\cdot$', '$<$', and '$\in$', the first five of which are governed by eight first-order "basic axioms", reminiscent of Peano's original system (Peano 1889), and providing a minimal keystone for Peano arithmetic,[15] whereas the sixth is governed by a second-order induction axiom and an unrestricted comprehension axiom-scheme, namely

$$\forall X[[0 \in X \wedge \forall n(n \in X \Rightarrow n+1 \in X)] \Rightarrow \forall n(n \in X)], \tag{1}$$

and the universal closure of

$$\left[\exists X \forall n \left[n \in X \Leftrightarrow \overset{\mathfrak{L}^{Z_2}}{\varphi}(n)\right]\right], \tag{2}$$

where '$\overset{\mathfrak{L}^{Z_2}}{\varphi}(n)$' stands for any formula of $\mathfrak{L}^{Z_2}$ in which 'X' does not occur freely.

This is a very strong theory, existentially speaking. However, it is also very specific. It is true that intending its variables to range over natural numbers and sets of them does not ipso facto restrict their range, and (2) is just the usual unrestricted comprehension scheme of full second-order logic extended to the whole $\mathfrak{L}^{Z_2}$. Nonetheless, Z_2 is such that its models only include items behaving as natural numbers and sets of them are ordinarily required to do, not only for their forming a progression, but also for their being linked to each other by the order, additive and multiplicative relations that are ordinarily required to hold for natural numbers. In other words, Z_2 is specifically about natural numbers, as bearing these relations, and about sets of them. Understanding it involves understanding the conditions characterising both of these relations and two sorts of items—such that the items of the first sort bear these relations to one another, whereas those of the second sort are sets of items of the first sort—and calls for the appropriates notions. In particular, understanding the conditions characterising the second sort of items calls for more

[15] Adding to these axioms a first-order axiom-scheme of induction, one gets the system Z_1, which provides a convenient version of Peano first-order arithmetic (Simpson 2009, pp. 7–8).

than the notion of a set of items of the first sort, because it also involves understanding how such a particular set is fixed through a condition expressed in $\mathfrak{L}^{Z_2}$.[16] All this demands extensive intellectual resources. Moreover, as it is typical of structural definitions, these resources are all required simultaneously and from the start, in order to gain epistemic access to these items and begin to work consciously with them, because these definitions determine *ex novo* the relevant items as *sui generis* items characterised by the relations they are required to bear.

As it is well known, the existential strength of Z_2 can be significantly reduced, while preserving much of its deductive strength, by restricting the comprehension axiom-scheme to formulas of an appropriate syntactical simplicity (cf. Simpson 2009 for a comprehensive study). This results in systems such as ACA_0, where comprehension is restricted to formulas containing no second-order quantifier, or RCA_0, where comprehension is restricted to Σ^0_1 and Π^0_1 formulas that are equivalent to one another and (1) is replaced by an axiom-scheme restricted to Σ^0_1 or Π^0_1 formulas. However, this does not lessen the intellectual resources required to understand the definition of natural numbers in Z_2. On the contrary, one could even argue that it increases these resources, because, on the one hand, restricting comprehension in a certain way does not result in limiting in the same way the complexity of the formulas involved in the relevant system, from the understanding of which depends the understanding of the system itself (to see it, note that the universal closure of an instance of a restricted comprehension axiom-scheme can have a greater syntactical complexity than the formulas to which this scheme is restricted),[17] and, on the other hand, understanding restricted comprehension involves understanding the criterion on which the restriction depends, which is, of course, not necessary to understand unrestricted comprehension. Moreover, though non-categorical (relatively both to their first and second order parts), ACA_0 and RCA_0 are, in a sense, even more specific than Z_2: the items they implicitly define are so specified as to be putatively suitable for supplying the building blocks to be used in the enterprise of recovering certain portions of "ordinary mathematics" (Simpson 2009, p. 1) on the basis of set-theoretic existential assumptions that are as weak as possible.

The situation with FA is quite different. Its language—let us call it '$\mathfrak{L}^{FA}$'—is much poorer than $\mathfrak{L}^{Z_2}$, because it reduces to the language $\mathfrak{L}^{L_2}$ of an appropriate system of full second-order logic L_2 including identity for terms and both monadic and dyadic predicate variables, supplemented by the single non-logical (functional) constant '#', introduced by (3), which, written in extenso and replacing schematic predicates with predicate variables, takes the following form:

[16]Cf. Footnote 18.

[17]The most evident case is that of a comprehension axiom-scheme restricted to formulas containing no second-order quantifier, as that involved in ACA_0: whatever such formula '$\varphi(n)$' might be, the syntactical complexity of '$\exists X \forall n[n \in X \Leftrightarrow \varphi(n)]$' is greater than that of this formula.

$$\forall F, G\left[\#F = \#G \Leftrightarrow \exists R\left[\begin{array}{l}[\forall x(F(x) \Rightarrow \exists! y(R(x,y) \wedge G(y)))] \wedge \\ [\forall y(G(y) \Rightarrow \exists! x(R(x,y) \wedge F(x)))]\end{array}\right]\right]. \quad (3)$$

Besides adding (3) to the axioms of L_2, to move from L_2 to FA, one must also extend comprehension, both for monadic and dyadic predicates, to formulas including '#', which results in replacing the comprehension axiom-schemes of L_2 with the universal closures of the following schemes:

$$\exists F \forall x\left(F(x) \Leftrightarrow {}^{\mathfrak{L}^{FA}}\varphi\,(x)\right) \quad \text{and} \quad \exists R \forall x, y\left(R(x,y) \Leftrightarrow {}^{\mathfrak{L}^{FA}}\varphi\,(x,y)\right) \quad (4)$$

where '${}^{\mathfrak{L}^{FA}}\varphi\,(x)$' and '${}^{\mathfrak{L}^{FA}}\varphi\,(x,y)$', respectively, stand for any formulas of $\mathfrak{L}^{FA}$ in which 'F' and 'R' do not occur freely.

This already shows that FA does not involve sets,[18] and, although dealing with cardinal numbers, is not specifically about them. Indeed, L_2 merely deals with objects, their properties, or first-level concepts, and dyadic relations among them, and (3), although suitable for implicitly defining cardinal numbers, does not allow one to prove that $\forall x \exists F[x = \#F]$, with the result that the models of FA can be populated by objects other than cardinal numbers. *A fortiori*, FA does not involve sets of natural numbers and it is not specifically about these numbers.

Natural numbers have, rather, to be explicitly defined within FA. By relying on comprehension and (3), one first defines 0 and the successor relation between cardinals:

$$\begin{array}{l}0 =_{df} \#[n{:}\, n \neq n] \\ \forall x, y[\mathcal{S}(x,y) \Leftrightarrow \exists F \exists z(F(z) \wedge y = \#F \wedge x = \#[n : F(n) \wedge n \neq z])]\end{array}. \quad (5)$$

[18]It should be noted that what matters here is not merely the way in which $\mathfrak{L}^{FA}$'s predicates are informally conceived, in particular by neo-logicists, as opposed to the way in which $\mathfrak{L}^{Z_2}$'s predicates are conceived. Indeed, intending the second-order variables of Z_2 to range over sets of elements of the range of the first-order ones, and the constant '$\in$' as designating the set-theoretic relation of membership is not mandatory. One could rather intend the second-order variables of Z_2 to range over the monadic properties of the elements of the range of the first-order ones, and consider that '$n \in X$' is nothing but a typographic variant of '$X(n)$' or 'Xn' (that is, merely an alternative way to predicate the property X of the individual n). What matters is rather the way in which predicates work in FA and Z_2, respectively. Focusing on the mere definition of natural numbers, the difference is not really significant, because what Z_2's predicates do in relation to this definition can, *mutatis mutandis*, also be done by FA's monadic predicates. The difference becomes, instead, quite significant in relation to the definition of real numbers within these theories (which I shall consider in the next section). Indeed, if second-order variables of Z_2 are taken to range over monadic properties of the elements of the range of the first-order ones, rather than over sets of these same elements, one can hardly be happy with a definition of real numbers as some particular items within the range of the former of these variables, as suggested by Simpson in relation to ACA_0 and RCA_0

Then, by relying on comprehension, again, one defines the strong ancestral $\mathcal{S}^*$ of $\mathcal{S}$, and defines natural numbers as those cardinals which are either 0 or bear the relation $\mathcal{S}^*$ with it:

$$\forall x[\mathcal{N}(x) \Leftrightarrow (0 = x \vee \mathcal{S}^*(0, x))], \tag{6}$$

where '$\mathcal{N}$' designates the property of being a natural number, of course. It follows that in FA, natural numbers are singled out among cardinal numbers, without exhausting the latter, because it is clear from (5) to (6) that there is at least one cardinal number, namely $\#\mathcal{N}$, which is not a natural number.

In many respects, this lack of specificity is not a virtue. However, it is also the symptom of FA's epistemic weakness, because it makes clear that what FA's definition of natural numbers does is not fixing these very numbers and the sets of them *ex novo* as *sui generis* items, but rather fixing, first, cardinal numbers within the putative range of the individual variables of $\mathfrak{Q}^{L_2}$, that is, among objects in general (conceived as the inhabitants of this range), and, next, singling out natural numbers among cardinal ones. Understanding this definition requires a considerable amount of intellectual resources. However, it seems to me that these are fewer, or at least more basic and more progressively appealed to, than those required to understand Z_2 as an implicit definition of natural numbers. Let us explains why.

First at all, it is clear that Z_2 includes, as does FA, a system of second-order logic with full comprehension, with the result that understanding Z_2 involves understanding this system, just as happens for FA. The system L_2 included in FA is both distinct and differently conceived than the one included in Z_2. Understanding the former calls for: the notions of an object, of a first-level concept (or property of an object),[19] and of a first-level dyadic relation; the notions of falling of an object under a (first-level) concept, and of a pair of objects bearing a certain (first-level) relation to each other; the notion of identity for objects; the notions of a variable ranging over objects, concepts, and dyadic relations, respectively; the notion of full comprehension both for first-level concepts and first-level dyadic relations; and all the notions generically involved in understanding a system of predicative logic. The difference with the system of second-order logic included in Z_2 depends on the fact that, in this last one, concepts are replaced by sets[20] and dyadic relations are avoided. Even so, the notion of a dyadic relation is required to understand the basic axioms of Z_2, and, once the notion of a second-order variable is at hand, moving from this to the notion of a variable ranging over dyadic relations does not require too much. Hence, if understanding L_2 requires more than understanding the system of second-order logic included in Z_2, the difference does not seem to be significant.

[19] Cf. Footnote 5.

[20] Cf. Footnote 18.

Second, all (3) does is appealing to a formula of $\mathfrak{Q}^{L_2}$ to fix some items—namely, cardinal numbers—in the putative range of its variables, by putting forward an identity condition for them. This definition is close to a structural one in some sense. Because the items forming the putative range of individual variables of L_2 lack sufficient characterisation to allow one to single out some of them among all of them, merely by specifying such a characterisation. Hence, one could say that (3) introduces cardinal numbers *ex novo* as *sui generis* items, as the axioms of Z_2 do for natural numbers and sets of them. Despite this, and whatever (3)'s existential import for objects might be, besides the notions mentioned above in connection with the understanding of L_2, understanding (3) merely calls for the notion of a many-one association between concepts and objects (better, between many concepts and a single object), which is needed to understand (3)'s left-hand side, plus the notion of the objects falling under a certain (first-level) concept being in bijection with those falling under another (first-level) concept, which is needed to understand the formula of $\mathfrak{Q}^{L_2}$ providing (3)'s right-hand side. It follows that, apart for what pertains to the notion of a many-one association between concepts and objects, HP's epistemic cost is structurally analogous to (i.e. is constituted in the same way as) the epistemic cost of any explicit definition of a sort of objects within any second-order theory.[21] Indeed, to provide such an explicit definition, one writes a formula of the appropriate language involving a single unbounded first-order variable, then appeals to comprehension to guarantee that there exists a property that an item belonging to the putative range of this variable has, or a set to which such an item belongs, if and only if it satisfies this formula, and finally introduces an appropriate monadic predicate constant to designate this property or set. Analogously, to introduce a predicate constant designating the property of being a cardinal number on the base of (3), one has to rely on comprehension (extended to '#'), to conclude that there exist a property that an object has if and only if it is uniquely associated with a certain concept F in such a way that it is the same object as any object equally associated with any concept G on condition that F and G satisfy the formula of $\mathfrak{Q}^{L_2}$ providing the right-hand side of (3), and then introduce such a predicate constant to designate this property.

Third, though (3)'s epistemic cost is minimal, understanding it allows one to have an epistemic access to cardinal numbers, and then begin to work consciously with them. Definitions (5), (6) result from this work. Both definitions (5) depend on (3) and comprehension, but are mutually independent, and also independent of any other definition, whereas definition (6) is, as such, independent of (3), but depends on comprehension and the two previous ones. It is just this definition that singles out natural numbers among cardinal ones. Hence, if, once having been defined through (3), cardinal numbers are taken to be particular items, natural ones are defined within FA by coding them with appropriate items. It follows that the definition of natural numbers coming with FA is neither structural nor close to a

[21]It should be noted that the notion of a many-one association between appropriate sorts of items is involved in any definition of whatsoever functional constant.

structural definition in the sense in which that of cardinal numbers via (3) is: it does not result from defining a structure, but rather from providing a system exemplifying it and formed by items singled out among items to which one has already had epistemic access. Moreover, it succeeds in this without appealing to any relation, or operation on natural numbers themselves, except for the relations $\mathcal{S}^*$ and $\mathcal{S}$, already defined on cardinal numbers. It follows that, besides the notions mentioned above in connection with the understanding of L_2 and (3), understanding the definitions (5) and (6) merely calls for the notions of a first-level concept under which no object falls, for that of a (first-level) concept under which falls all the objects falling under another (first-level) concept plus a single one, and for that of the strong ancestral of a first-level dyadic relation.

Once one has obtained an epistemic access to natural numbers as defined though (3), (5) and (6), by appealing to the notions just mentioned, one can conscientiously work on them, in turn, namely define appropriate relations and operations on them, and prove that, under these relations and operations, they exemplify the same structure as that defined by the axioms of Z_2 (which is what is usually called 'Frege's theorem'). Hence, it seems to me that so defining natural numbers is not only epistemically more economic than doing it within Z_2, but also epistemically more economic than doing it within any version of Peano second-order arithmetic, even if this does not include axioms for addition, multiplication, and order. Indeed, any definition coming with any version of Peano second-order arithmetic is structural, and allows one to have access to natural numbers only insofar as at least one appropriate relation is defined on them. Moreover, in any such definition, zero is not singled out among items to which one has an independent epistemic access, but is simply posited by fixing some conditions it has to meet. This should be enough to conclude that the definition of natural numbers coming with FA is epistemically economic, in my sense.

Epistemic Economy and the Definitions of Real Numbers

The epistemic advantage of appropriate abstraction principles over other forms of definitions becomes even clearer in the case of the definition of real numbers.

As it is well known, there are several ways to define real numbers within second-order arithmetic. First I consider a definition within Z_2, which follows Cantor's classical approach, and I then compare it with three definitions within FA (or, strictly speaking, within appropriate extensions of it). Even though the former merely depends on supplying the axioms of Z_2 with some appropriate explicit definitions, whereas the latter all depend on adding new appropriate abstraction principles to FA's axioms, I argue that the epistemic cost of the former is comparable to that of two of the latter, although the third is epistemically (more) economic.

Defining Real Numbers Within Z_2

The first definition is proposed by Simpson (2009, § I.4) as a definition with ACA_0, which can also be repeated, with a minor change in its last step, within RCA_0. Its purpose is to show how to define real numbers within second-order arithmetic while considerably weakening the existential set-theoretic assumptions on which Z_2 depends. This is a crucial fact, underlying the whole program of reverse mathematics. However, for my present purpose, it can be ignored, and the definition can be merely taken as a definition within Z_2. There is certainly room for arguing that the mere possibility of achieving this definition within ACA_0, and one close to it within RCA_0, shows that this definition actually depends on much weaker existential assumption than those on which, not only Z_2, but also FA, and consequently any definition of real numbers involving this latter theory, depend. This still does not mean, in my view, that any definition of real numbers within FA has a greater epistemic cost than Simpson's within ACA_0 and RCA_0. This contrast between the advantage of Simpson's definitions in terms of existential parsimony and the lack of benefit in terms of epistemic economy is just part of the point I make. To this end, there is no essential difference between immersing Simpson's definition within ACA_0 or RCA_0 and immersing it within the whole Z_2.[22]

This definition proceeds in four steps. In the first, the set $\mathbb{N}$ of natural numbers is defined as the unique set X such that $\forall n(n \in X)$, which is licensed by comprehension and by the specificity of Z_2. In the second step, integer numbers are coded with elements of an appropriate subset $\mathbb{N}_{\mathbb{Z}}$ of $\mathbb{N}$ and order, addition, and multiplication are defined on them. Details left aside and supposing that ζ is whatever natural number, this results in coding the integers $+\zeta$ and $-\zeta$ with $\zeta^2+\zeta$ and ζ^2, respectively. In the third step, rational numbers are coded with the elements of another appropriate subset $\mathbb{N}_{\mathbb{Q}}$ of $\mathbb{N}$ and order, addition, and multiplication are defined on them, too. Details left aside, anew, and supposing that ζ and ϑ are whatever pair of coprime natural numbers, the second of which is positive, the rational numbers $+\frac{\zeta}{\vartheta}$ and $-\frac{\zeta}{\vartheta}$ are, respectively, coded with $\left(\zeta^2+\zeta+\vartheta^2+\vartheta\right)^2+\zeta^2+\zeta$ and $\left(\zeta^2+\vartheta^2+\vartheta\right)^2+\zeta^2$. In the fourth step, sequences of rational numbers are coded with appropriate subsets of $\mathbb{N}$, namely subsets

[22]Once this definition is immersed within the whole Z_2, the items it defines—namely (the items re-casting) real numbers within Z_2—provably have many properties that the items it defines when it is immersed within ACA_0 or RCA_0—namely (the items re-casting) real numbers within ACA_0or RCA_0—do not provably have. The crux of reverse mathematics (to which Simpson 2009 is entirely devoted) is just which properties of the former items are preserved once real numbers are defined within weaker sub-systems of Z_2, like ACA_0 or RCA_0. However, of course, this is not a matter I can consider here.

meeting an appropriate condition, and real numbers by some of these subsets, namely those suitable for coding Cauchy sequences of rational numbers, in agreement with Cantor's definition (1872, pp. 123–124).[23]

This short description is enough to show that, according to this definition, integer, rational and real numbers are not introduced *ex novo* as *sui generis* objects, but rather singled out among items previously defined—namely natural numbers for integer and rational ones, and sets of them, for real ones—through explicit definitions that are independent of any operation or relation on these very numbers. First, integer numbers are singled out among natural ones, by appealing to addition, multiplication and order on the latter. Addition, multiplication and order are then defined on integer numbers, and appealed to in order to single out rational numbers among them.[24] Finally, addition, multiplication, order, and absolute value are defined on rational numbers, and appealed to in order to single out real ones among sets of them. A definition of addition, multiplication and order on real numbers only comes into play at this point, together with the proof that these numbers behave with respect to them so as to comply with the required structural conditions.

It is then clear that the intellectual resources needed to understand the whole definition are not involved simultaneously from the start. Understanding its different steps rather involves understanding a limited number of formulas at once, which are used to single out the relevant items among others to which one already has epistemic access. Still, understanding the definition of real numbers involves understanding the previous definitions of natural, integer and rational ones and of the relevant operations and relations on them, and then calls for all the resources that are needed to understand these previous definitions. Furthermore, it also involves the understanding of the notion of a Cauchy sequence of rational numbers, and the use of (implicitly defined)[25] sets of natural ones for coding other numbers (that is, of second-order items to code first-order ones), and then calls for the corresponding resources. Even though this is a large amount of resources, it does not include those which are needed to understand the definitions of any operation and relation on real numbers themselves. These definitions come later and are not

[23]It is easy to see the essential difference between this fourth step and the three previous ones: whereas in these three steps, the sets $\mathbb{N}$, $\mathbb{N}_{\mathbb{Z}}$, and $\mathbb{N}_{\mathbb{Q}}$ are explicitly defined, the subsets of $\mathbb{N}$ coding real numbers cannot be explicitly defined, in turn, and it is, a fortiori, impossible to define anything working as the set of real numbers. All that one can do is fixing a condition that a subset of $\mathbb{N}$ has to met in order to code a single real number.

[24]In fact, integer numbers could be singled out among natural ones by appealing only to addition and multiplication on the latter, by merely stipulating that the former numbers are coded by those of the latter ones which are equal to $\zeta^2+\zeta$ or ζ^2, for some natural number ζ. In this way, no justification could be offered for this choice. Analogously, rational numbers could be directly singled out among natural numbers, by only appealing, again, to addition and multiplication on the latter, by merely stipulating that the former numbers are coded by those of the latter ones which are equal to $\left(\zeta^2+\zeta+\vartheta^2+\vartheta\right)^2+\zeta^2+\zeta$ or $\left(\zeta^2+\vartheta^2+\vartheta\right)^2+\zeta^2$, for some pair of coprime natural numbers ζ and ϑ, the second of which is positive. In this case also, no justification could be offered for this choice.

[25]Cf. Footnote 23.

needed in order to have an epistemic access to these numbers. This is enough to show that the definitions of real numbers, as well as those of integer and rational ones, are not structural. If, once having been implicitly defined through the axioms of Z_2 as places in a structure, natural numbers and sets of them are taken to be particular items, the definitions of integer, rational and real numbers all consist of coding these numbers with some of these particular items, forming instances of other structures.

Defining Real Numbers Within FA

I now return to FA. Nothing would prevent one from rephrasing Simpson's definition of integer and rational numbers within this theory, so as to code them with natural numbers appropriately singled out. Rendering Simpson's definition of real numbers within FA would be more problematic, because FA provides no definition of sets of natural numbers, and real ones could therefore not be singled out among these sets. One could, at most, state a certain condition that a property of natural numbers should meet in order either to code real numbers directly or to belong to a range of (second-order) variables entering an abstraction principle implicitly defining these numbers as the values of a functional operator taking the elements of this range as its arguments. Insofar as, in FA's abstractionist setting, the former option would be at odds with the idea that real numbers are objects, similar to natural, integer and rational ones, the latter would certainly be preferable. However, going for it would, in turn, be more convoluted (and certainly epistemically less economic) than taking advantage of FA's lack of specificity and appealing to an appropriate abstraction principle to directly fix real numbers within the putative range of FA's individual variables—namely among objects, and especially among other ones than cardinal numbers, thus replicating, *mutatis mutandis*, the same move already made to define cardinal numbers through FA. This is the option chosen by the three definitions of real numbers within FA, which I now consider.

To this end, I take, for short, '$\forall_\Phi$' and '$\exists_\Phi$'—where 'Φ' stands for a monadic predicate constant—to designate, respectively, the universal and the existential quantifiers restricted either to items that have Φ or to their properties.[26]

Shapiro's Rephrasing of Dedekind's Definition

The first definition is Shapiro's rephrasing of Dedekind's within FA (Shapiro 2000, pp. 338–340 and Dedekind 1872). Similar to Simpson's, Shapiro's definition includes previous definitions of integer and rational numbers, but, unlike in

[26]In other terms, I shall take '$\forall_\Phi x[\phi]$', '$\forall_\Phi X[\phi]$', '$\exists_\Phi x[\phi]$', and '$\exists_\Phi X[\phi]$' to abbreviate '$\forall x[\Phi(x) \Rightarrow \phi]$', '$\forall X[\forall x[X(x) \Rightarrow \Phi(x)] \Rightarrow \phi]$', '$\exists x[\Phi(x) \wedge \phi]$', and '$\exists X[\forall x[X(x) \Rightarrow \Phi(x)] \wedge \phi]$' respectively.

Simpson's, these numbers are not singled out among natural ones. Similar to real ones, they are rather fixed within the putative range of the individual variables of $\mathfrak{Q}^{\mathrm{FA}}$ through two abstraction principles, one for integers, the other for rational numbers. I suggest that we call 'FDRA' (for 'Frege-Dedekind Real Arithmetic'), the extension of FA that is obtained by adding to its axioms these two principles together with what is required to define real numbers, and by extending comprehension to formulas including the functional constants so introduced.

The abstraction principle used to define integer numbers is as follows:

$$\forall_{\mathcal{N}} x, x', y, y' [\mathrm{INT}(x, y) = \mathrm{INT}(x', y') \Leftrightarrow x + y' = x' + y], \tag{7}$$

where 'INT' is the functional constant introduced by this principle, and '$+$' designates the addition on natural numbers. To be more precise, this principle merely provides an implicit definition of appropriate pairs of natural numbers; integer ones are then defined by coding them with these pairs: if p and q are whatever pair of natural numbers, $\mathrm{INT}(p, q)$ is taken, within FDRA, as an integer number. This results from explicitly defining a predicate constant—let us say '$\mathcal{Z}$'—designating the property of being such a number. Appealing to this constant, after having explicitly defined the multiplication on integer numbers, licenses a new abstraction principle:

$$\forall_{\mathcal{Z}} x, x', y, y' \left[\begin{array}{c} \mathrm{QUOT}(x, y) = \mathrm{QUOT}(x', y') \\ \Leftrightarrow \left[\begin{array}{c} [y = 0_{\mathcal{Z}} \wedge y' = 0_{\mathcal{Z}}] \vee \\ [y \neq 0_{\mathcal{Z}} \wedge y' \neq 0_{\mathcal{Z}} \wedge x \cdot_{\mathcal{Z}} y' = x' \cdot_{\mathcal{Z}} y] \end{array} \right] \end{array} \right], \tag{8}$$

where 'QUOT' is the functional constant introduced by this principle, '$0_{\mathcal{Z}}$' abbreviates 'INT$(0, 0)$'—the integer zero—, and '$\cdot_{\mathcal{Z}}$' designates the multiplication on integer numbers. This principle implicitly defines appropriate pairs of integer numbers. Rational numbers are then defined as those of these pairs whose second element is not $0_{\mathcal{Z}}$: if u and v are a pair of integer numbers and $v \neq 0_{\mathcal{Z}}$, $\mathrm{QUOT}(u, v)$ is taken, within FDRA, as an integer number. This also results from an explicit definition introducing a predicate constant—let us say '$\mathcal{Q}$'—designating the property of being an integer number. Appealing to this constant, after having explicitly defined the order relation on rational numbers, allows one to define explicitly the (second-order) relation that a property of rational numbers has with such a number if the latter is an upper bounded of the former:

$$\forall_{\mathcal{Q}} F \forall_{\mathcal{Q}} x [F \trianglelefteq x \Leftrightarrow \forall_{\mathcal{Q}} y [F(y) \Rightarrow y \leq_{\mathcal{Q}} x]], \tag{9}$$

where '$\leq_{\mathcal{Q}}$' designates the order relation on rational numbers. Finally comes a third abstraction principle:

$$\forall_{\mathcal{Q}} F, G [\mathrm{CUT}(F) = \mathrm{CUT}(G) \Leftrightarrow \forall_{\mathcal{Q}} x (F \trianglelefteq x \Leftrightarrow G \trianglelefteq x)], \tag{10}$$

where 'CUT' is again a functional constant introduced by this principle. This new principle implicitly defines cuts on properties of rational numbers. Real ones are finally defined by coding them with appropriate such cuts: if P is a property of rational numbers, CUT(P) is to be taken as (coding) a real number if and only if P is instantiated and has an upper bound, that is, it is such that $\exists_{\mathcal{Q}} x, y[P(x) \wedge P \trianglelefteq y]$.

This definition is quite natural from the neo-logicist perspective, but it differs structurally from Simpson's one within Z_2 only by appealing to cuts of rational numbers, rather than to (sets of natural numbers coding) Cauchy sequences of these same numbers, and by replacing the corresponding explicit definitions with abstraction principles working as new axioms. One could think that this latter circumstance makes Shapiro's definition epistemically more costly than Simpson's. However, I do not think it is so.

First, FA provides a basis for the former definition which is epistemically weaker than Z_2, or any sub-system of it, even though it remains that the mere definition of natural numbers within FA is not enough to license the subsequent definition of integer, rational, and real ones within this same theory, and has to be supplied by the definition of addition on natural numbers and of multiplication and order on rational ones.

Second, as it happens with (3), these principles appeal to a formula of $\mathfrak{Q}^{\mathrm{FA}}$—appropriately extended, in the case of (8), and (10)—to fix some items in the putative range of FA's individual variables by putting forward an identity condition for these items. Hence, understanding these principles does not call for more than understanding the explicit definition of a sort of object depending on the introduction of a monadic predicate constant apt to designate the property that these items are required to have, or the set they form. I have already made a similar point with respect to (3). For (7), (8), and (10), the point is even clearer. To illustrate it with an example, compare Simpson's explicit definition of rational numbers within Z_2 with (7). To get the former one focuses on the open formula '$\exists m[h = m^2 \vee h = m^2 + m]$', relies on comprehension to conclude that $\exists X \forall h[h \in X \Leftrightarrow \exists m[h = m^2 \vee h = m^2 + m]]$, and takes rational numbers to be the elements of the set of natural numbers which have the property satisfying this last condition. To define rational numbers through (7) within FA, one states this principle, relies on comprehension appropriately extended to 'INT' to conclude that $\exists F \forall z[F(z) \Leftrightarrow \exists_{\mathcal{N}} x, y[z = \mathrm{INT}(x, y)]]$ and takes rational numbers to be objects that have the property satisfying this last condition.

Finally, similar to Simpson's, Shapiro's definition defines real numbers by providing an example of the relevant structure, rather than defining this structure as such.

It seems, then, that the two definitions have analogous epistemic costs, or, even, that Shapiro's one is epistemically more economic.

Hale's Rephrasing of Frege's Definition

The second definition I consider results from adapting to FA Hale's rephrasing of Frege's definition of domains of magnitudes and of his plan to get a definition of real numbers as ratios on such a domain (Hale 2000, pp. 105–113; Frege 1893, part III, §§ II 55–245; Simons 1987; Dummett 1991, Chap. 22; Schirn 2013). According to Frege, real numbers originate in measuring magnitudes, and they have to be defined as ratios of them, rather than as arithmetical items (as happens in Cantor's and Dedekind's definitions). However, though independent of natural numbers, his definition of domains of magnitudes takes place in the same (inconsistent) system of logic in which he also defines these numbers, and could be rephrased within L_2 (appropriately strengthened). The successive definition of real numbers as ratios on such a domain can, then, easily be achieved by resorting to natural numbers as a useful auxiliary tool, but can also be freed from any recourse to these numbers (although then becoming a little bit more convoluted). In the former case, the definition would depend on an appropriate extension of FA; in the latter, it would depend on an extension of L_2, both different from, and independent of, FA. In both cases, one should, however, pair the definition with an existence proof for domains of magnitudes, which Frege merely outlines informally. His idea is to obtain such a domain by starting from natural numbers, whose existence is taken for granted. Formally rendering his indications then results, quite naturally, in a construction within FA. Faced with this situation, Hale provides an informal, algebraically shaped, definition of domains of magnitudes appealing to natural numbers (without specifying the way they are defined), then suggests an existence proof of these domains, deviating from Frege's indications, but still based on natural numbers and on the admission of their existence, and finally defines ratios of magnitudes by resorting to natural numbers. The simplest way to adapt his definition to FA, though openly departing from Frege's conception, is by directly defining positive real numbers as ratios on the specific domains of magnitudes drawn from natural numbers during the existence proof.[27] This is the plan I follow.

Hale first defines "normal quantitative domains" as pairs $\langle \mathbf{Q}, + \rangle$ where: $\mathbf{Q}$ is non-empty and closed under +; + is associative, commutative and such that if $\mathbf{p}$ and $\mathbf{q}$ are distinct elements of $\mathbf{Q}$, there is another element $\mathbf{r}$ of $\mathbf{Q}$ for which either $\mathbf{p} = \mathbf{q} + \mathbf{r}$ or $\mathbf{q} = \mathbf{p} + \mathbf{r}$; for any $\mathbf{p}$ and $\mathbf{q}$ in $\mathbf{Q}$, there is an $\mathbf{r}$ in $\mathbf{Q}$ and a positive natural number n such that $\mathbf{q} + \mathbf{r} = n\,\mathbf{p} = \underbrace{\mathbf{p} + \mathbf{p} + \cdots + \mathbf{p}}_{n \text{ times}}$. If a strict order $<$ is defined on $\mathbf{Q}$ by stipulating that $\mathbf{p} < \mathbf{q}$ if and only if there an $\mathbf{r}$ in $\mathbf{Q}$ such that $\mathbf{q} = \mathbf{p} + \mathbf{r}$, then $<$ is a total order on $\mathbf{Q}$, and it becomes easy to show that $\langle \mathbf{Q}, + \rangle$ meets the Archimedean condition (and includes only positive elements). Ratios on normal quantitative domains are then defined by an abstraction principle rephrasing

[27] Hale's domains of magnitudes, unlike Frege's, only include, as we see below, positive elements, with the result that only positive real numbers can be defined as ratios on them. Non-positive ones are, then, to be defined by extension.

definition V.5 of Euclid's *Elements*, stating that, if $\langle \mathbf{Q}, + \rangle$ and $\langle \mathbf{Q}^*, + \rangle$ are normal quantitative domains (not necessarily distinct), and $\mathbf{p}$, $\mathbf{q}$ are in $\mathbf{Q}$ and $\mathbf{p}^*$, $\mathbf{q}^*$ in $\mathbf{Q}^*$, then:

$$\mathrm{RAT}(\mathbf{p}, \mathbf{q}) = \mathrm{RAT}(\mathbf{p}^*, \mathbf{q}^*) \Leftrightarrow \forall_{NN^+} h, k \left[h\mathbf{p} \lesseqgtr k\mathbf{q} \Leftrightarrow h\mathbf{p}^* \lesseqgtr k\mathbf{q}^* \right], \qquad (11)$$

where 'NN^+' designates the property of being a positive natural number (however defined). Next, domains of magnitudes (which Hale rather calls 'complete normal quantitative domains') are defined as normal quantitative domains that meet the fourth-proportional condition (for any $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ in $\mathbf{Q}$, there is a $\mathbf{s}$ in $\mathbf{Q}$ itself, such that $\mathrm{RAT}(\mathbf{p}, \mathbf{q}) = \mathrm{RAT}(\mathbf{r}, \mathbf{s})$) and are Dedekind-complete.

To prove that domains of magnitudes exist (i.e. that at least one such domain exists), Hale takes, as I have said above, the existence of natural numbers for granted, and relies on them to obtain such a domain by following a path analogous to that involved in Shapiro's foregoing definition. He begins by observing that positive natural numbers form, together with the addition on them, a normal quantitative domain, let us say $\langle \mathbf{N}^+, + \rangle$. One can then define ratios on it through (11), by taking $\mathbf{Q}$ and $\mathbf{Q}^*$ to coincide with each other and with $\mathbf{N}^+$, and easily verify that these ratios, together with an appropriate addition on them, form, in turn, a normal quantitative domain meeting the fourth-proportional condition, let us say $\left\langle \mathbf{R}^{\mathbf{N}^+}, + \right\rangle$. This allows cuts to be defined on $\mathbf{R}^{\mathbb{N}^+}$ through a new abstraction principle, which is nothing but a restriction of Frege's Basic Law V:

$$\mathrm{CUT}(P) = \mathrm{CUT}(Q) \Leftrightarrow \forall x_{R^{\mathbf{N}^+}} [P(x) \Leftrightarrow Q(x)] \qquad (12)$$

where '$R^{\mathbf{N}^+}$' designates the property of being an element of $\mathbf{R}^{\mathbf{N}^+}$, and P and Q are whatever properties of the elements of $\mathbf{R}^{\mathbf{N}^+}$ that are non-empty, non-total, downward closed, and upward unbounded. These cuts form the required domain of magnitudes, let us say $\left\langle \mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}, + \right\rangle$.

The structure of domains of magnitude is categorical. Hence, any ratio on $\mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}$ is identical with a ratio on any other domain of magnitudes (if any). This allows one to define positive real numbers by coding them with ratios on $\mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}$, because these ratios are the same as those on any other such domain (if any).

Similar to Shapiro's, Hale's definition of real numbers includes two stages: first some items are defined by appealing to an appropriate abstraction principle, then these items are appealed to in order to define real numbers. There are, however, important differences between the two definitions. The most evident is that the first stage of Hale's definition is not only more laborious than that of Shapiro's, but it is also informal and algebraic in spirit: it requires an existence proof and does not specify how the elements of a normal quantitative domain (and, then, also of a domain of magnitudes) are fixed, with the result that the only connection between this proof and the definition depends on verifying afterwards that the systems

constructed meet the relevant condition. This does not make Hale's definition of real numbers structural. This definition does not define positive real numbers as places in a structure defined as such. However, it is not intended to code these numbers with the items defined in the first stage of the definition, namely with ratios on domains of magnitudes. In agreement with Frege's purpose, it is instead intended to identify these numbers as such, to disclose their ultimate nature, namely that of being these very ratios (which might be described in different ways, according to the particular domain of magnitudes that are chosen, provided that different such domain exist: taking these ratios, and then positive real numbers, to be ratios on $\mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}$ is, indeed, nothing but a convenient way to name or describe them). This identification is, however, quite questionable, certainly more questionable than the neo-logicist idea (also inspired by Frege) that natural numbers are just the values of the function # singled out by (6). It is much more unquestionable to admit that ratios on domains of magnitudes merely code positive real numbers.

Together with the possibility of defining (or describing) these ratios as ratios on $\mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}$, this suggests freeing Hale's definition from the definition of domains of magnitudes, and incorporating it within FA. Indeed, all that matters for the definition to be appropriate, at least from a purely mathematical point of view, is that the ratios on $\mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}$ provide suitable codes for real numbers, independently of their being regarded as ratios of magnitudes. It remains, however, that $\mathbf{C}^{\mathbf{R}^{\mathbf{N}^+}}$ cannot be defined, as such, within FA, because this cannot but be a set, and no set can be defined within FA. One has rather to replace it, as well as $\mathbf{N}^+$ and $\mathbf{R}^{\mathbf{N}^+}$, with appropriate properties. One first defines the property $\mathcal{N}^+$ of being a positive natural number, and rephrases (11), by replacing '**p**', '**q**', '$\mathbf{p}^*$' and '$\mathbf{q}^*$' with first-order variables bounded by '$\forall_{\mathcal{N}^+}$', '$\forall_{NN^+}$' with '$\forall_{\mathcal{N}^+}$' again, and 'RAT' with '$\mathrm{RAT}_{\mathcal{N}^+}$'. This allows one to define explicitly the property $\mathcal{R}^{\mathcal{N}^+}$ as being a value of the function $\mathrm{RAT}_{\mathcal{N}^+}$, and rephrase (12), by replacing '$\forall_{R^{\mathbf{N}^+}}$' with '$\forall_{\mathcal{R}^{\mathcal{N}^+}}$', '*P*' and '*Q*' with second-order variables bounded by '$\forall_{\mathcal{R}^{\mathcal{N}^+}}$' again, and 'CUT' with '$\mathrm{CUT}_{\mathcal{R}^{\mathcal{N}^+}}$'. One explicitly defines, then, the property $\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}$ as being a value of the function $\mathrm{CUT}_{\mathcal{R}^{\mathcal{N}^+}}$ for the argument given by a non-empty, non-total, downward closed, and upward unbounded property of the items having the property $\mathcal{R}^{\mathcal{N}^+}$, rephrasing (11), again, by replacing '**p**', '**q**', '$\mathbf{p}^*$' and '$\mathbf{q}^*$' with first-order variables bounded by '$\forall_{\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}}$', and 'RAT' with '$\mathrm{RAT}_{\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}}$' (by assuming that addition and strict order have been appropriately defined on the items having the property $\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}$). Finally, one codes positive real numbers with the values of the function $\mathrm{RAT}_{\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}}$. This results in supplementing FA with three abstraction principles, each of which is followed by a corresponding explicit definition. In fact, the second rephrasing of (11) could be avoided by directly coding positive real numbers with the items having the property $\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}$. In both cases, a further step, involving either a new

abstraction principle, or a suitable explicit definition, is required to move from positive real numbers to real numbers *tout court*.

Whatever opinion one might have of Frege's requirement that real numbers are to be defined as ratios of magnitudes, and, consequently, of the philosophical appropriateness of Hale's original definition of real numbers, it is doubtless that its epistemic cost is high, because its understanding involves the understanding of the definition of the structure of domains of magnitudes, and of the existence proof for domains of magnitudes coming with the definition of $\mathbf{C}^{\mathbf{R}^{\mathbb{N}^+}}$. However, if Hale's definition of real numbers is adapted to FA in the way suggested above, it becomes very close to Shapiro's, except for possibly involving an eliminable definition of ratios of cuts on ratios of positive natural numbers (i.e. a definition of ratios on items having the property $\mathcal{C}^{\mathcal{R}^{\mathcal{N}^+}}$) and for obtaining non-positive real numbers in the end, rather than defining integer numbers in the beginning. On the one hand, the first rephrasing of (11) replaces (8), limitatively to positive natural numbers, and comes quite close to it, in fact, because, if 'x', 'x'', 'y' and 'y'' range over these numbers, the two conditions that $\forall_{\mathcal{N}^+} h, k\left[hx \lesseqgtr kx' \Leftrightarrow hy \lesseqgtr ky'\right]$ and that $x \cdot_{\mathcal{Z}} y' = x' \cdot_{\mathcal{Z}} y$ are provably equivalent within FA. On the other hand, the rephrasing of (12) replaces, limitatively to positive rational numbers, both (10) and the restriction to cuts of instantiated and upper bounded properties of these numbers. Hence, except for the subtle differences that one might discern among the respective epistemic costs of these two pairs of axioms and among that of the stipulation required to obtain non-positive real numbers and that of (8), the epistemic cost of the two definitions is the same, and then either analogous to that of Simpson's definition or smaller than it.

Real Numbers as Bicimal Pairs

Even though it openly departs from Frege's indication as well, the last definition I consider is suggested by one of his ideas, namely by his outline of the existence proof of domains of magnitudes (to which Hale does not conform, as I have said above).

Frege's heuristic suggestion goes as follows. Look at Cauchy's series of the form $\sum_{i=0}^{\infty} \lambda_i \frac{1}{2^i}$, where λ_0 is a natural number, and λ_i $(i = 1, 2, \ldots)$ are either 0 or 1 but are not constantly 0 after a certain value of i. These series are in bijection both with positive real numbers (because any such series converges to such a number, any such number is the limit of a such a series, and distinct series converge to distinct numbers and vice versa), and with all the pairs $\langle \lambda_0, \mathfrak{S} \rangle$, where $\mathfrak{S}$ is the (infinite) set of positive natural numbers i such that $\lambda_i = 1$ (because, given any such series, one can get such a pair, and vice versa). There is thus a bijection between positive real numbers and these pairs. Now, looking at these pairs as such, one can define an internal addition on them, let us say $\uplus$, without any consideration of real numbers, and then use it as a basis to define a family of permutations among these pairs: to

any such pair α, one associates the permutation $\uplus_\alpha$ such that, for any two other pairs β and γ, $\uplus_\alpha(\beta, \gamma)$ if and only if $\beta = \alpha \uplus \gamma$. There is then also a bijection between these pairs and these permutations, and then between the latter and positive real numbers. The idea is, then, to show that the value-ranges of these permutations, together with those of their converses and of the identity permutation, form a domain of magnitudes under the operation of composition (defined on permutations, then transferred to their values-ranges).

As I have said, Frege's purpose is to define real numbers as ratios of magnitudes. Hence, the heuristic interest of noting that the relevant permutations are in bijection with positive real numbers is not that of showing that the latter can be coded with the former, or better with their value-ranges (and non-positive real numbers with the converses of these permutations together with the identity permutations, or with their value-ranges), but rather that of showing that there are enough relevant permutations for their value-ranges to form a domain of magnitudes. Still, Frege's outline naturally suggests to define real numbers, or, at least, positive ones, by coding them with pairs such as $\langle \lambda_0, \mathfrak{S} \rangle$, or with some appropriate *alias* of them. Hilbert and Bernays do something close to this in their *Grundlagen der Mathematik* (Hilbert and Bernays 1934, vol. II, supplement IV, § C). My suggestion is to render this idea within FA by generalising it straight away to non-positive real numbers. In particular, I suggest that FA be extended, so as to define pairs such as $\langle p, P \rangle$—where p is a natural number and P an infinite (that is, instantiated and upward unbounded) property of natural numbers—and to code real numbers with these pairs.

To this end, it is enough to supplement FA with a single abstraction principle and a single explicit definition, and to extend comprehension to formulas including the constant introduced by this principle. I suggest calling 'FRA', for 'Frege Real Arithmetic', the extension of FA obtained in this way.

The abstraction principle is the following:

$$\forall_{\mathcal{N}} x, y \forall_{\mathcal{N}} X, Y[\langle x, X \rangle = \langle y, Y \rangle \Leftrightarrow [x = y \wedge \forall_{\mathcal{N}} z[X(z) \Leftrightarrow Y(z)]]], \tag{13}$$

where '$\langle -, - \rangle$ ' is a dyadic functional constant introduced by this principle.

The explicit definition is required to impose that the properties entering the relevant pairs are infinite. It is as follows:

$$\forall x[\mathcal{B}(x) \Leftrightarrow \exists_{\mathcal{N}} y \exists_{\mathcal{N}} Y[x = \langle y, Y \rangle \wedge \forall_{\mathcal{N}} z \exists_{\mathcal{N}} w[\mathcal{S}^*(z, w) \wedge Y(w)]]]. \tag{14}$$

This definition introduces the monadic predicate constant '$\mathcal{B}$', designating the property of being a value of the function $\langle -, - \rangle$ when its second argument is an infinite property. I call the items having this property 'bicimal pairs', and I suggest real numbers are coded with them.

Axiom (13) and definition (14) are enough to complete the definition. However, in order to show that this definition is appropriate, it is also necessary to define addition, multiplication and strict order on bicimal pairs, so that they behave with respect to these operation and this relation as real numbers are required to behave. Of

course, this is the case of any definition of these numbers depending on coding them with appropriate items defined beforehand, and then also of the three definitions considered above. However, in this last case, the definitions of these operations and this relation are less immediate than in the previous ones, though it should be clear that providing these definitions would be nothing but a question of logical routine. I have no space here to detail this routine. All I can say is how I intend to distinguish positive from non-positive real numbers from the very beginning, that is, by relying neither on these last definitions, nor on the definition of the real zero.

The basic idea is as follows. Let q be any natural number, and Q and $\tilde{Q}$ any two properties of natural numbers such that $Q(0), \neg\tilde{Q}(0)$ and $Q(n) \Leftrightarrow \tilde{Q}(n)$ for any positive natural number n. According to (13) and (14), the pairs $\langle q, Q\rangle$ and $\langle q, \tilde{Q}\rangle$ are distinct from each other, even though they correspond to the same pair $\langle q, \mathfrak{S}\rangle$, where $\mathfrak{S}$ is such that i belongs to it if and only if it is a positive natural number such that $Q(i)$. Hence, bicimal pairs are not in bijection with positive real numbers. However, they are in bijection with real numbers *tout court*, because, if $\langle q, Q\rangle$ is associated with the positive real $\rho = q + \sum_{i=1}^{\infty} \lambda_i \frac{1}{2^i}$, where $\lambda_i = 1$ if and only if $Q(i)$, $\langle q, \tilde{Q}\rangle$ can be associated with the non-positive real $\rho - 2q - 1 = \sum_{i=0}^{\infty} \lambda_i \frac{1}{2^i} - q - 1$. This suggests taking a bicimal pair $\langle p, P\rangle$ to be positive if P is such that $P(0)$, and to be non-positive if P is such that $\neg P(0)$, and coding positive and non-positive real numbers with positive and non-positive bicimal pairs, respectively.

This is rendered by explicitly defining the two properties $\mathcal{B}^+$ and $\mathcal{B}^{0/-}$ as follows:

$$\forall x\big[\mathcal{B}^+(x) \Leftrightarrow [\mathcal{B}(x) \wedge \exists_{\mathcal{N}} y \exists_{\mathcal{N}} Y[x = \langle y, Y\rangle \wedge Y(0)]]\big], \tag{15}$$

$$\forall x\Big[\mathcal{B}^{0/-}(x) \Leftrightarrow [\mathcal{B}(x) \wedge \exists_{\mathcal{N}} y \exists_{\mathcal{N}} Y[x = \langle y, Y\rangle \wedge \neg Y(0)]]\Big], \tag{16}$$

and by coding positive and non-positive real numbers, respectively, with the items (belonging to the putative range of the individual variables of $\mathfrak{L}^{L_2}$) that have these properties.

Once this is done, one can take the bicimal pair $\langle 0, \mathcal{N}^+\rangle$—which is clearly such that $\mathcal{B}^{0/-}(\langle 0, \mathcal{N}^+\rangle)$—to code the real zero, and define addition and strict order on bicimal pairs so that a positive real ρ and a negative one $-\rho$ are, respectively, coded with the bicimal pairs $\left\langle \lfloor p \rfloor_\rho, P\right\rangle$ and $\left\langle \left|\lfloor p \rfloor\right|_\rho, \tilde{P}\right\rangle$, where: $\lfloor p \rfloor_\rho$ is the greatest natural number strictly smaller than ρ; P is such that $P(0)$ and that $P(i)$ if and only if $\sum_{i=1}^{\infty} \lambda_i \frac{1}{2^i} = \rho - \lfloor p \rfloor_\rho$ is such that $\lambda_i = 1$; $\left|\lfloor p \rfloor\right|_\rho$ is the greatest natural number smaller or equal to ρ; and $\tilde{P}$ is such that $\neg\tilde{P}(0)$ and that $\tilde{P}(i)$ if and only if $\sum_{i=0}^{\infty} \lambda_i \frac{1}{2^i} = \rho - \left|\lfloor p \rfloor\right|_\rho$ is such that $\lambda_i = 0$.

Defining real numbersthis way allows the task to be achieved without relying on any previous definition of integer and rational numbers. Moreover, just as positive and non-positive real numbers are discriminated, and the real zero is identified

afterwards, once all real numbers have been defined at once and the same time by supplementing FA with (13) and (14)—, integer and rational numbers can be easily defined later by discerning them among real ones. To this end, it is enough to take integer numbers to be the real ones which are coded with bicimal pairs whose second element is either $\mathcal{N}$ or $\mathcal{N}^+$, and rational numbers to be the real numbers that are coded with bicimal pairs whose second element is a periodic property of natural numbers, that is, a property P of these numbers such that $\exists_{\mathcal{N}}x, y\forall_{\mathcal{N}}z[\mathcal{S}^*(x,z) \Rightarrow [P(z) \Leftrightarrow P(z+y)]]$. Once this is done, it is also quite simple to discern positive from non-positive integers and rational numbers: an integer number is positive if and only if it is coded with a bicimal pair whose second element is $\mathcal{N}$, and it is non-positive if and only if it is coded with a bicimal pair whose second element is $\mathcal{N}^+$; a rational number is positive if and only if it is coded with a bicimal pair whose second element is a periodic property of natural numbers that is enjoyed by 0, and it is non-positive if and only if it is coded with a bicimal pair whose second element is a periodic property of natural numbers that is not enjoyed by 0. Finally, the integer and the rational zero simply coincide with the real one.

It does not only follows that understanding this definition of real numbers is independent of understanding any definition of integer and rational numbers, but also that understanding the definition of real numbers as such—which merely consist of (13) and (14)—provides most of what is required to understand several other subsequent definitions, namely that of the real zero, those of positive and negative real numbers, those of integer and rational numbers, and, among them, of positive and non-positive such numbers, and of the integer and the rational zero. Each of these definitions is not only quite simple but it is also fully independent of the definition of any operation and relation on integer, rational, and real numbers themselves. Moreover, both the definitions of positive and non-positive real numbers and of the real zero, and those of integer numbers, and, among them, of positive and non-positive ones, and of the integer zero, are independent of the definition of any operation on natural numbers themselves, as well as of any relation on them other than the two relations $\mathcal{S}(x,z)$ and $\mathcal{S}^*(x,z)$, which already enter the definition of these last numbers. A previous definition of addition on natural numbers is only required to define rational numbers, and, together with multiplication, to define the usual operations and relations on real numbers (and, consequently on integer and rational ones), and to prove that these numbers behave with respect to these operations and relations as they are required to do.[28]

This makes clear that understanding this definition of real numbers, as well as all those subsequent ones, requires only a small amount of intellectual resources. Indeed, apart from the notions that are needed to understand the relevant system of second-order logic, and the notion of a many-one association between pairs such as

[28]This proof is quite convoluted, but combinatorial in spirit, and epistemically quite economic, because, elementary arithmetic on natural numbers being taken for granted, it does not involve much more than propositional logic applied to predicate (second-order) formulas.

$\langle p, P \rangle$ and objects, which is needed to understand the left-right side of (13), understanding the definitions of real numbers, and, among them, of positive and non-positive such numbers, and of the real zero, merely calls for: the notions of a natural number and of a property of natural numbers—which includes, in this setting, those of cardinal numbers, of the successor relation between these numbers, and of its strong ancestral—; the notions of a variable ranging over these numbers and over their properties, respectively; the notion of the natural zero, as distinguished from any other natural number; the notions of the identity and the strict order relation among natural numbers; the notion of a natural number having a property; and, finally, the notions of a property of natural numbers being enjoyed or not by a certain natural number, and being enjoyed by exactly the same natural numbers as another such property. Besides these notions, to understand the definitions of integer numbers, and, among them, of positive and non-positive such numbers, and of the integer zero, only the notion of a positive natural numbers—i.e. of a natural number other than zero—is called for. Finally, to understand the definitions of rational numbers, and, among them, of positive and non-positive such numbers, and of the rational zero, it is enough to add the notion of addition on natural numbers. These resources are not only much fewer than those required to understand the three previous definition of real numbers, but they are also very basic. Noting this seems to me enough to conclude that the definition of real numbers coming with FRA is epistemically economic.

Conclusions

The previous comparative assessments of two definitions of natural numbers and four definitions of real ones were intended to show that: (i) alternative formal definitions of the same mathematical items can be discriminated according to their respective epistemic cost, and a choice among them can be made so as to prefer the one whose epistemic cost is the smallest, which (provided that the comparison involves a comprehensive and representative enough sample of formal definitions of the relevant items) I suggest to qualify as epistemically economic *tout court* (though, of course, other choices can be legitimately made on the basis of some other criteria); (ii) in the cases under consideration, i.e. in relation to (second-order) definitions of natural and real numbers, the appeal to appropriate abstraction principles, within appropriate settings, namely to (3) and to (13), allows one to obtain epistemically economic definitions; (iii) this does not depend on the existential strength of these principles, and thus, a fortiori, on their existential parsimony or ontological neutrality (which, by the way, I do not think they benefit from), showing that existential parsimony or ontological neutrality and epistemic economy are independent virtues. From all this, it follows that the epistemic economy of the definitions of natural and real numbers, respectively, coming with FA and FRA provides a possible reason to favour FA over other versions of

(second-order) arithmetic, both as such, and as a base for real analysis, a reason which is independent of FA's existential strength.

My considerations leave, however, many related issues open. I wish to address two of them in conclusion.

Analyticity

The first issue concerns the relation between epistemic economy and analyticity.

Famously, Dedekind opened the preface to the first edition of *Was sind und was sollen die Zahlen* (Dedekind 1888) by declaring that arithmetic is "the simplest science", namely a "part of logic", and that it is so insofar as "the number-concept [is][…] an immediate result [*Ausfluss*] from the pure laws of thought", because "numbers are free creations of the human mind […][which] serve as a means of apprehending more easily and more sharply the difference of things", and "counting an aggregate or number of things" depends on "the ability of the mind to relate things to things, to let a thing correspond to a thing, or to represent a thing by a thing, an ability without which no thinking is possible" (Dedekind 1901, p. 14, with a slight modification).[29] H. Benis-Sinaceur takes this to mean that arithmetic is a part of logic because "numbers […] are rooted in the constitution of the mind or, as Dedekind writes to Keferstein (February 27, 1890), they are 'subsumed under more general notions and under activities […] of the understanding […] without which no thinking is possible' " (Benis-Sinaceur 2015, § 1.3.1; Sinaceur 1974, p. 272; von Heijenoort 1879, p. 100). The same point is also made clear in this other passage drawn from the fragment *Zum Zahlbegriff* (which Benis-Sinaceur quotes only partially: (Benis-Sinaceur 2015, § 1.6): "Of all the resources that the human mind [is endowed with] for relieving its life, that is, [for fulfilling its] task, none is so effective and so inseparably connected with its inner nature as the concept of number […], because every thinking man, even if he is not clearly aware of that, is a number-man, an arithmetician" (Dugac 1976, Appendix LVIII, p. 315).[30]

Dedekind's logicism then seems to consist of the thesis that the resources we use to count, and, more generally, to deal with natural numbers, are just the same as, or part of, those we use to think *tout court*, and in the identification of logic with the intellectual activity exercising these resources. Even though Dedekind and Frege

[29]Though he generically speaks of numbers (*Zahlen*), Dedekind's claims seem to be directly referred to natural numbers. Still, he also seems to consider that his views on these numbers extend to any other sorts of numbers, insofar as theories of the latter come from an extension of the theory of the former (or arithmetic, as usually intended). This is made clear by the parenthesis in the following claim: " In speaking of arithmetic (algebra, analysis) as a part of logic I mean to imply that I consider the number-concept [*Zahlbegriff*] entirely independent of the notions or intuitions of space and time, that I consider it an immediate result from the laws of thought" (Dedekind 1901, p. 14).

[30]I thanks Emmylou Haffner for drawing my attention to this passage.

have often been associated as two partisans of logicism—generally presented as the thesis that arithmetic can be reduced to logic by adopting an appropriate definition of natural numbers—this point of view is quite different from Frege's (emphasising this difference is the main purpose of Benis-Sinaceur's paper just mentioned; on this matter, cf. also the Introduction of Benis-Sinaceur et. al 2015). One would still misunderstand Frege's point if one regarded it as the mere affirmation that arithmetic can be recovered within his logical system. It was, indeed, part of this point that this system is epistemically basic, insofar as it pertains to the basic components of any thought.[31] On the other hand, Dedekind's purpose was not merely to account for the intrinsic features of arithmetic; it was also to prescribe the right definition of natural numbers: "upon this unique and therefore absolutely indispensable foundation […] must, in my judgement, the whole science of numbers be established", he also writes in the mentioned preface, just after the passage quoted above (Dedekind 1901, p. 14). Doubtless, Frege's logic and thought are not activities, and, for him, natural numbers are certainly not "free creations of the human mind", as for Dedekind. Still, this crucial difference should not obscure a more fundamental agreement: that the logicality of arithmetic depends on its generality, which results, in turn, from its dealing with the building blocks of any other possible science (be it an exercise of human reason, as for Dedekind, or a system of truths, as for Frege), and that the definition of natural numbers has to conform to this.

It is then tempting to associate Dedekind's regarding the logicality of arithmetic as its involving the (or same as the) basic resources of thinking with Frege's view that arithmetical truths are analytic insofar as their proof only depends on "logical laws and definitions", and to suggest that a definition of some mathematical items is analytic insofar as its understanding only calls for logical resources. If this were admitted, there would also be room for assenting to Dedekind's view that the creative import of a definition does not preclude its being logical in nature. Exegetically speaking, one could doubt that Dedekind's regarding numbers as human mind's creations amounts to ascribing an ontological import for objects to whatsoever definition of them. Still, this would be independent from being ready to

[31]Look at the two following quotes, from the *Grundlagen* (Frege 1884, § 14; 1953, p. 21), and from a coeval paper ("Über Formale Theorien der Arithmetik", Frege 1967, pp. 103–111, especially p. 103; English translation in Frege 1984, pp. 112–121, especially p. 112), respectively:

> The basis of arithmetic lies deeper, it seems, than that of any of the empirical sciences, and even than that of geometry. The truths of arithmetic govern all that is numerable. This is the widest domain of all; for to it belongs 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.

> As a matter of fact, we can count about everything that can be an object of thought: the ideal as well as the real, concepts as well as objects, temporal as well as spatial entities, events as well as bodies, methods as well as theorems; even numbers can in their turn be counted. What is required is really no more than a certain sharpness of delimitation, a certain logical completeness.

consent to the idea that admitting that a definition of natural numbers has such an import should not prevent one from considering that its understanding only calls for logical resources, and that it is then analytic in the tentative sense just mentioned. Of course, it would still remain to explain what it means for the understanding of a definition to call only for logical resources. This could be difficult to do in a precise way, but the previous considerations about the definitions of natural and real numbers coming with FA and FRA suggest that this construal of the notion of analyticity leaves room to argue that natural numbers, and possibly also real ones, admit an analytic definition. This would result in a vindication of Frege's views, although in a philosophical setting essentially different both from that of Frege's himself, and from that of the neo-logicist.

To defend the strong neo-logicist analyticity thesis, one could maintain, then, that understanding (3) and (5), (6) only calls for logical resources by arguing as follows. If it were admitted that second-order logic is logic, or, more precisely, that L_2 is a genuine system of logic, it would be natural to maintain that understanding the right-hand side of (3) only calls for logical resources. Hence, if it were also admitted that understanding the universal closure of a double implication '$\boldsymbol{S}(f) \Leftrightarrow \boldsymbol{S}$', introducing the new constant 'f' (non-occurring in '$\boldsymbol{S}$'), only calls for logical resources if this is the case for understanding '$\boldsymbol{S}$', it would follow that understanding the whole (3) only calls for logical resources. Furthermore, if it were equally admitted that understanding an explicit definition in $\mathfrak{L}^{L_2} + \{f\}$ only calls for logical resources if this is the case for understanding the universal closure of '$\boldsymbol{S}(f) \Leftrightarrow \boldsymbol{S}$', it would also follow that understanding (5), (6) only calls for logical resources, as well. It would then be enough to be ready to make the three mentioned admissions to conclude that the definition of a natural number coming with FA is not only epistemically economic, as I have argued above, but it is also analytic in the foregoing sense.

If all this were conceded, it should also be possible to go ahead in a similar vein and argue that understanding (13)–(16) only calls for logical resources, in turn, with the result that the definition of real numbers coming with FRA would also be analytic. However, why should it not be possible to fashion a similar argument supporting the claim that this is also the case of Shapiro's definitions of real numbers, as well as of that expounded above, deriving from adapting Hale's one to FA? In the face of this option, one could adopt three different attitudes.

First, one could look for reasons to block the argument in relation to these two latter definitions, although admitting it in relation to the former one, in order to conclude that, whereas the former definition is analytic, the latter two aren't. For this purpose, one could advance, for example, that there is a relevant difference in the epistemic cost of the universal closure of a double implication '$\boldsymbol{S}(f) \Leftrightarrow \boldsymbol{S}$' introducing the new constant 'f', according to whether '$\boldsymbol{S}$' is a formula of $\mathfrak{L}^{L_2}$, or '$\boldsymbol{S}$' includes some constants that do not belong to $\mathfrak{L}^{L_2}$ (and this independently of whether this universal closure is unrestricted or admits a restriction involving some predicate constants), with the result that from admitting that understanding such a

universal closure in the former case only calls for logical resources does not entail that this is also the case for understanding it in the latter case.[32]

Second, one could question that an argument similar to the previous one, supporting the conclusion that the definition of natural numbers coming with FA is analytic, applies to FRA, and then maintain that neither the definition of real numbers coming with FRA, nor Shapiro's, nor the adaptation of Hale's to FA are analytic. For this purpose, one could argue, for example, that the mere fact that an abstraction principle incorporates a universal quantification restricted by using a predicate constant entails that the understanding of this principle calls for more than only logical resources. More radically, one could also question that understanding an explicit definition in $\mathfrak{Q}^{L_2} + \{f\}$ only calls for logical resources if this is the case for understanding the definition of 'f', so as also to block, in this way, the previous argument bringing the conclusion that the definition of natural numbers coming with FA is analytic.

Finally, one could accept that Shapiro's definitions of real numbers and the adaptation of Hale's one to FA are both analytic, after all, just like to the one coming with FRA, although conceding that distinct analytic definitions of the same items could have significantly different epistemic costs, with the result that only one of them is epistemically economic.

Exemplarist Definitions

I now turn to the second question.

At the end of the paper in which he presents his definition, Shapiro touches on Heck's distinction between interpreting arithmetic "in some analytically true theory" and showing that "the truths of arithmetic, as we ordinarily understand them, are analytic", and he wonders whether the "cuts on bounded, instantiated properties of rational numbers [as defined by (10)] are the real numbers that we all know and love?" (Shapiro 2000, pp. 360–361; Heck 1997, p. 596). This last question is, as such, independent of the admission that FA, and possibly also FDRA, are analytic, and can be repeated for any definition of some mathematical items depending on coding these items with others previously defined, that is, as one could say, for

[32]To see the point, remark, firstly, that $\mathfrak{Q}^{L_2}$ includes no predicate constant, so that a restriction involving some predicate constant cannot be stated in $\mathfrak{Q}^{L_2}$. Remark, then, that in (13), the restricted quantifier '$\forall_{\mathcal{N}}z$' can be equivalently replaced by an unrestricted one (its entering this principle is only motivated by easiness of understanding). Remark, finally the difference between the open formula '$x = y \wedge \forall z\,[X(z) \Leftrightarrow Y(z)]$', providing the right-hand side of (13), and the other open formulas '$x + y' = x' + y$', '$[y = 0_{\mathcal{Z}} \wedge y' = 0_{\mathcal{Z}}] \vee [y \neq 0_{\mathcal{Z}} \wedge y' \neq 0_{\mathcal{Z}} \wedge x \cdot_{\mathcal{Z}} y' = x' \cdot_{\mathcal{Z}} y]$', '$\forall_{\mathcal{Q}}x(F{\trianglelefteq}x \Leftrightarrow G{\trianglelefteq}x)$', '$\forall_{\mathcal{N}^+}h, k\left[hx \lesseqgtr ky \Leftrightarrow hx^* \lesseqgtr ky^*\right]$' and '$\forall_{\mathcal{R}^{\mathcal{N}^+}}[P(x) \Leftrightarrow Q(x)]$', providing the right-hand sides of the abstraction principles involved in Shapiro's definitions of real numbers, and in the adaptation of Hale's to FA: whereas the first of these formulas is a formula of $\mathfrak{Q}^{L_2}$, this is so for none of the others.

short, for any exemplarist definition. By only considering the definitions of natural and real numbers, respectively, coming with FA and FRA, the question becomes: should we regard these theories as genuine theories of natural and real numbers, as we ordinarily understand them, or, merely, as theories within which arithmetic and real analysis are interpretable?

As I see it, the question is not whether FA and FRA truly define, respectively, natural and real numbers, that is, if the items that, within these theories, are taken to be these numbers are actually these very numbers. So conceived, the question makes little or no sense, it seems to me, unless one looks at the matter from a hyper-realist ontological perspective, which I neither share, nor believe could be adopted by default. In my view, the question is rather whether the definitions of these numbers, respectively, coming with FA and FDRA have what is essential to the nature we attribute to these numbers built into themselves. Of course, if one maintains that there is nothing essential to this nature besides these number's meeting the relevant structural conditions, then this question also makes little sense. However, maintaining this is far from mandatory: there is room to consider that what is essential to this nature also depends—or even depends only—on the place we attribute to these numbers in our mathematical knowledge as a whole (for example, on the mutual relations between natural and real numbers that follow from our way to conceive them), or, more generally, in the whole system of our knowledge.

In this perspective, the relevant question with respect to FA is whether we should consider that taking natural numbers to be trademarks of the cardinality of finite concepts reflects what is essential to their nature. In this form, the question has often been discussed, for example considering whether defining natural numbers through FA meets the application constraint. There is then no need to come back to it, here. I confine myself to observe that nothing ensures, in general, that a definition whose understanding calls for less and/or more basic resources than others or even an analytic definition in the foregoing sense has what is essential in the nature we attribute to the relevant items built into it. Hence, there is no general reason to think that FA's being epistemically economic goes together with its having what is essential in the nature we attribute to natural numbers built into it. Its possibly having both virtues would then be a supplementary epistemic advantage that this definition would have over alternatives, because this would let it show that what is essential in the nature we attribute to natural numbers can be grasped by appealing to few and quite basic resources. Moreover, if it happened that these resources could be considered as merely logical, this would mean that logical resources are enough to grasp what is essential in the nature we attribute to natural numbers, which could be taken as a proper way to state a logicist thesis for someone who, along with me, merely considers mathematics as a result of our intellectual activity.

It is then relevant to wonder whether something similar could also be said of FRA, supposing that it be regarded as providing an epistemically economic definition of real numbers, or even an analytic definition of them in the foregoing sense.

Against the idea that the definition of real numbers depending on FRA has all what is essential in the nature we attribute to these numbers built into it, one could

observe that the explicit definitions making bicimal pairs behave as real numbers are guided by a previous understanding of the relevant structure, together with the admission that real numbers exemplify this structure, and that this suggests that taking real numbers to be (coded with) bicimal pairs might be considered appropriate only if these numbers are independently conceived to exemplify this structure. Now, it is certainly not built into the definition of bicimal pairs that they exemplify this structure: this can only be verified a posteriori with respect to the definition itself. Hence, if what is essential to the nature we ascribe to real numbers includes their exemplifying this structure, defining these numbers as (coded with) bicimal pairs cannot have all what is essential to this nature built into it.

This is a plausible argument, but there are reasons to resist it. One could indeed argue that what is essential to the nature we ascribe to real numbers does not include their exemplifying the relevant structure, but only their possibly doing so, should appropriate operations and relations be defined on them. To make an analogous point with respect to natural numbers, one could argue that what is essential to the nature we ascribe to them is not their behaving with respect to order, addition and multiplication as they do, but their being so that this relation and these operations can be defined on them so that they behave in this way. Arguing that natural numbers are essentially cardinal numbers (or numbers of concepts) rather than elements of a progression is, after all, a way to make this point. Could not one make an analogous point for real numbers? If this were conceded, it would be relevant to note that having an epistemic access to bicimal pairs is independent of having an epistemic access to the relevant structure. Indeed, this would leave room to maintain that the mere resources needed to understand (13) and (14) are enough to grasp what is essential to the nature we attribute to real numbers. Surely. However, leaving room to maintain that this is so is still not the same as providing reasons for it. Hence the question remains: are the resources needed to understand (13) and (14) enough to grasp what is essential to the nature we attribute to real numbers? To argue that this is not so, and so contest that the definition of real numbers depending on FRA have what is essential to this nature built into it, one could observe that it is a fact that our ordinary understanding of real numbers does not involve bicimal pairs. The following considerations should, however, be enough to overcome this objection and support a positive answer to the question.

Insofar as the notions of limit, convergence, continuity and cut are in no way appealed to in the definition of real numbers depending on FRA (though proving that bicimal pair form a complete group cannot but require appealing to some of them), this definition suggests that there is a way to understand the key notions of real analysis—which are certainly involved in our ordinary understanding of real numbers—that results from our acquaintance with an instance of the structure of real numbers, rather than the other way round. Hence, one could maintain that taking real numbers to be (coded with) bicimal pairs allows us to have an epistemic access to these numbers in such a way that we can, then, and only then, get our ordinary understanding of them through working on them by exploiting the possibilities embodied in the very nature of these pairs. If it were conceded that what is essential to the nature we ascribe to real numbers merely includes their possibly

exemplifying the relevant structure if some appropriate operations and relations are defined on them, it would follow that this would already be implicitly built into these numbers being coded with bicimal pairs, and it would only be a question of making it explicit.

In the paper mentioned above, Benis-Sinaceur argues that Dedekind's original definition of real numbers, unlike Cantor's, "is not based on the concepts of limit and convergence", and that this definition rather "shows how to derive the concept of limit, and thus the usual theorems of real analysis, from the purely arithmetical definition of the concept of real numbers" (Benis-Sinaceur 2015, § 1.1). This suggests that Dedekind's definition already makes clear that understanding some key notions of real analysis can result from our acquaintance with an instance of the real numbers structure, rather that with this structure itself. Nonetheless, understanding Dedekind's definition involves understanding a previous definition of rational numbers and of cuts on (and then sets of) them, which is not the case of the definition depending on FRA. *Mutatis mutandis*, the same also happens for Shapiro's definition and for the adaptation of Hale's one to FA. Hence, if it were conceded that the definition depending on FRA has what is essential to the nature we ascribe to real numbers built into it, this would show that there is a way to grasp what is essential to this nature that is epistemically more economic than the way displayed by these alternative definitions.

However, there is more. Insofar as FRA is an extension of FA, and defining real numbers through FRA depends on defining natural numbers through FA, so defining real numbers brings forward an idea of these last numbers as reifications of properties of cardinal numbers combinatorially steered. This is not the idea of real numbers that Frege attached to them, and is certainly not the idea that arises from looking at their applications in geometry and science. However, it is, it seems to me, a rather natural view of them both arithmetically speaking, and from a logicist perspective. Were it admitted that this very idea embodies what is essential to the nature we ascribe to real numbers—which is not only a possibility that the previous considerations leaves open, but also a very natural admission, if it were conceded that we essentially conceive real numbers as being numbers in the same sense as that in which natural numbers are so—, there would be no doubt that the definition of real numbers depending on FRA has what is essential to the nature we ascribe to these numbers built into it. And were it also admitted that this definition is analytic, in the foregoing sense, this would result in a version of the logicist thesis for real numbers.

However, even if all this were admitted, and this version of logicism were endorsed, this would not entail, in my view, that FRA provides the right or the best definition of real numbers. In my view, philosophy of mathematics should not aim at deciding which is the best way of defining certain mathematical items, or of structuring or founding mathematics or certain branches of it. It should rather aim (among other things) at identifying different philosophical virtues of different ways of defining mathematical items, and structuring or founding some branches of mathematics. My purpose was to isolate one of these virtues, namely epistemic economy, and to show that appealing to appropriate abstraction principles is suitable to obtain definitions that enjoy this virtue.

References

Banks, E. C. (2004). The philosophical roots of Ernst Mach's economy of thought. *Synthese, 109*, 23–53.

Benis-Sinaceur, H. (2015). Is Dedekind a logicist? Why does such a question arise? In Benis-Sinaceur, Panza, & Sandu (Eds.) *Reflections on Frege's and Dedekind's logicism: Functions and generality of logic*, pp. 1–57.

Benis-Sinaceur, H., Panza, M., & Sandu, G. (2015). *Reflections on Frege's and Dedekind's logicism: Functions and generality of logic*. Chem, Heidelberg, New York, Dordrecht, London: Springer.

Boolos, G. (1997). Is Hume's principle analytic? In R. J. Heck Jr. (Ed.), *Logic, language and thought: Essays in honour of Michael Dummett* (pp. 245–262). Oxford: Clarendon Press. Also in Boolos (1998), pp. 301–314 [I refer to this latter edition].

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

Cadet, M., & Panza, M. (2015). The logical system of Frege's G*rundgesetze*: A rational reconstruction. *Manuscrito*, *38*(1), 5–94.

Cantor, G. (1872). Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen. *Mathematische Annalen, 5*, 123–132.

Dedekind, R. (1872). *Stetigkeit und irrationale Zahlen*. F. Vieweg und Sohn, Braunschweig.

Dedekind, R. (1888). *Was sind und was sollen die Zahlen*. F. Vieweg und Sohn, Braunschweig.

Dedekind, R. (1901). *Essays on the theory of numbers*. Chicago: Open Court P. C.

Demopoulos, W. (Ed.). (1995). *Frege's philosophy of mathematics*. Cambridge, MA and London: Harvard University Press.

Dugac, P. (1976). *Richard Dedekind et les fondements des mathématiques*. Paris: Vrin.

Dummett, M. (1991). *Frege: Philosophy of mathematics*. London: Duckworth.

Frege, G. (1803/1903). *Grundgesetze der Arithmetik*. H. Pohle, Jena, 1893 (vol. 1), 1903 (vol. 2).

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

Frege, G. (1953). *The foundations of arithmetic* (J. L. Austin, Trans.). Oxford: Blackwell.

Frege, G. (1967). *Kleine Schriften*. G. Olms, Hildesheim, 1967. Herausgegeben von I. Angelelli.

Frege, G. (1984). *Collected papers on mathematics, logic, and philosophy* (B. McGuinness, Ed.). Oxford, New York: Basil Blackwell.

Hale, B. (2000). Reals by abstraction. *Philosophia Mathematica* (3rd series, Vol. 8, pp. 100–123).

Hale, B., & Wright, C. (2000). Implicit definition and the *a priori*. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 286–319). Oxford: Clarendon Press. Also in Hale & Wright (2001), pp.117–150.

Hale, B., & Wright, C. (2001). *The reason's proper study* (pp. 307–332). Oxford: Clarendon Press.

Hale, B., & Wright, C. (2009). Focus restored: Comments on John MacFarlane. *Synthese, 170*, 457–482.

Heck jr., R. J. (1997). Finitude and Hume's principle. *Journal of Philosophical Logic*, *26*, 589–617.

von Heijenoort, J. (Ed.). (1967). *From Frege to Gödel. A source book in mathematical logic, 1879–1931*. Cambridge, MA: Harvard University Press.

Hilbert, D., & Bernays, P. (1934–1939). *Grundlagen der Mathematik*. Berlin, Heidelberg, New York: Springer (2 Vols., 2nd ed., in 1968–1970).

Husserl, E. (1900). *Logische Untersuchungen. Erster Teil: Prolegomena zur reinen Logik*. M. Niemeyer, Halle a.d.S., 1900^1, 1913^2.

Husserl, E. (2001). *Logical investigations* (2 Vols., J. N. Findaly, Trans., with a new Preface by M. Dummett and edited with a new Introduction by D. Moran). London and New York: Routledge.

MacFarlane, J. (2009). Double vision: Two questions about the neo-Fregean program. *Synthese, 170*, 443–456.

Mach, E. (1883). *Die Mechanik in inher Entwickelung. Historisch-Kritisch Dargestellt*. F. A. Brockhaus, Leipzig (Many successive editions).
Peano, G. (1889). *Arithmetices principia nova methodo exposita*. Bocca, Torino. English translation in *Selected works of Giuseppe Peano*, H. C. Kennedy (Trans., and Ed., 1973, pp. 101–134). London: G. Allen & Unwin LTD.
Schirn, M. (2013). Frege's approach to the foundations of analysis (1874–1903). *History and Philosophy of Logic, 34*(3), 266–292.
Shapiro, S. (2000). Frege meets Dedekind: A neologicist treatment of real analysis. *Notre Dame Journal of Formal Logic, 41*(4), 335–364.
Shapiro, S., & Weir, A. (2000). 'Neo-logicist' logic is not epistemically innocent. *Philosophia Mathematica* 3rd series, *8*, 160–189.
Simons, P. (1987). Frege's theory of real numbers. *History and Philosophy of Logic*, *8*, 25–44. Also in Demopoulos (1995), pp. 358–383.
Simpson, S. G. (2009). *Subsystems of second order arithmetic* (2nd ed.). Cambridge, New York: Cambridge University Press (1st ed., 1999. Berlin: Springer).
Sinaceur, M. A. (1974). L'infini et les nombres. Commentaires de R. Dedekind à "Zahlen". La correspondance avec Keferstein. *Revue d'Histoire des Siences*, *27*, 251–278. Including a transcription (with French translation) of Dedekind's letter to Keferstein of February 2nd 1890.
Wright, C. (1999). Is Hume's principle analytic? *Notre Dame Journal of Formal Logic*, *40*, 6–30. Also in Hale & Wright (2001), pp. 307–332.

Author Biography

Marco Panza is research director at the CNRS, attached to the IHPST, Paris (CNRS and Univ. of Paris 1 Panthéon-Sorbonne). He is the author of several volumes and papers in the domain of history and philosophy of mathematics, especially mathematics at the early modern age and enlightenment, and the discussion on logicism and platonism in the contemporary philosophy of mathematics. His books include: *Newton et les origines de l'*analyse (*1664–1666*), Blanchard, Paris, 2005; and (with A. Sereni), *Plato's Problem. An Historical Introduction to Mathematical Platonism*, Palgrave, Basingstoke (UK), 2013.

Torn by Reason: Łukasiewicz on the Principle of Contradiction

Graham Priest

> *It is surprising how strongly certain opinions can persist within the sciences that are not only incorrectly formulated and without justification but which are plainly false—most likely, as I believe, because what has been declared in the past is repeated uncritically again and again.*
>
> Jan Łukasiewicz (Heine 2013, p. 125).

Introduction: A Seminal Book

In 1910, the young Jan Łukasiewicz published his ground-breaking *On the Principle of Contradiction in Aristotle*.[1] About two and a half millennia earlier, Aristotle had launched an attack on those who would violate the Principle of Non-Contradiction (PNC, or the Principle of Contradiction, as Łukasiewicz calls it). That attack established the principle as orthodox in Western Philosophy—in a way that perhaps no other philosophical claim has ever been so entrenched. The only Western philosopher who balked seriously was Hegel (and perhaps some of his intellectual descendants). Łukasiewicz subjected the Aristotelian arguments to a detailed, penetrating attack, and put the question of the cogency of the PNC on the table for 20th Century philosophy.

He was well aware of what he was doing. He says in the introduction to the book[2]:

[1] Łukasiewicz (1910a). He was 21 at the time.

[2] Heine (2013), p. 81. Quotations and page references in what follows are to this translation. All italics are original.

G. Priest (✉)
University of Melbourne, Melbourne, Australia
e-mail: priest.graham@gmail.com

G. Priest
CUNY Graduate Center, New York, USA

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_18

> There are two moments in the history of philosophy in which disputes over the principle of contradiction excited the minds of an age—one is connected with Aristotle's name—the other with Hegel. Aristotle formulated the principle of contradiction as the highest law of thought and being. He pursued everyone who would not recognize the principle with stubborn polemics in which, at times, anger and annoyance find a voice: Antisthenes and his school, Eristics from Megara, followers of Heraclitus, students of Protagoras. Aristotle won the fight. And whether it was the persuasive force of his arguments or the correctness of the position he defended—for centuries no one dared to contradict this highest of laws. Only Hegel allowed the convictions that had been buried by Aristotle to come back to life and instructed us to believe that reality is simultaneously rational and contradictory.

With considerable prescience, Łukasiewicz foresaw a third moment in the historical debate, one which would make use of the newly emerging symbolic logic[3]:

> If I am not mistaken, the *third* moment in the history of the principle is approaching now, a moment that will remedy old short-comings… [A] time has come in which logicians are beginning to review… [formal logical principles] and to dedicate themselves to those investigations that had not been considered by Hegel.

Indeed, he saw himself and his book as a harbinger of this moment. Before the third moment can be pursued properly, however[4]:

> one has to first return to Aristotle himself; some unresolved problems (related to the principle), which nowadays have been forgotten, need to be brought to mind and new investigations should then connect to them. I want to convince the reader that this principle is not as unshakable as one might expect with its general acceptance. I want to show that it presents a thesis which demands proof, and that despite the Stagirite's words… this proof can be found.

The Book's Two Halves

In the same year in which he published the book, Łukasiewicz published a paper in German, '*Über den Satz von Widerspruch bei Aristoteles*',[5] summarising the first half of his book. It is in this half that Aristotle is clinically dissected. The paper was—somewhat belatedly—translated into English after some 60 years; and its contents are now well appreciated by English-speaking philosophers. The book itself has not been translated into English until recently.[6] The present paper is an analysis of the contents of its second half.

In this part of the book we find Łukasiewicz giving his own views of the PNC, including the proof referred to in the last quote of the previous section. Łukasiewicz' categorical statement of the existence of a proof might lead one to expect him to have a similar categorical attitude to the PNC itself; but this is not what we find.

[3]Introduction, pp. 84–5.

[4]Introduction, pp. 86–7.

[5]Łukasiewicz (1910b).

[6]Heine (2013).

Łukasiewicz, as we shall see, is badly torn. It is true that he comes down on the side of the PNC eventually, but the journey is a tortured one, and the conclusion disappointingly lame. The material, however, shows us an acute mind wrestling with a principle it would really like to believe, in *despite* of the considerations it marshals. It also presents a fascinating window on a period in the history of logic, a century yore, when the new symbolic logic, the logical paradoxes, the thought of Meinong, and the thought of Hegel, delivered a heady and stimulating cocktail.

The Demolition of Aristotle

The critique of Aristotle in the first half of the book is clinical and devastating. It is also, as I have said, well known. Before we turn to the second half of the book, however, I want to note four of its aspects for future reference.

1. Aristotle's arguments for the PNC in the *Metaphysics* comprise one long argument, and then a series of about half a dozen brief arguments. The long argument is tangled, torturous, and how best to understand it is not at all clear. According to Łukasiewicz, it works, at best, only if applied to claims of the form Pa, where P is essentially predicated of a[7]:

> Even if… [the argument] were correct, it would prove the principle of contradiction only for a narrow range of objects: it would merely concern the essence of things, but not accidental properties.

Moreover, Łukasiewicz argues[8]:

> Everything appears to be in favour of the view that Aristotle limited the significance of the principle of contradiction to substantial being.

Łukasiewicz' claim is certainly contentious.[9] I cite these passages to show that he was well aware that one might hold that the principle has only limited validity. Indeed, in discussing the difference between the PNC and the principle of double negation, he himself appears to state that the PNC does not have universal validity. Discussing squaring the circle (a mathematical impossibility), he says[10]:

> There are instances in which the principle of double negation is true and the principle of contradiction is not applicable or, put simply, where it is false… Whoever has studied geometry will without doubt understand what this is: "A square constructed with the aid of compass and ruler, whose surface area is identical to the surface area of a circle that has radius 1"… Such a square—let's designate it simply by Q—is a contradictory object… Q *has* S [sides that are expressable by an algebraic number] and simultaneously *does not have* S. It is precisely because of this that Q is a contradictory object…

[7]Chapter XI, p. 153.

[8]Chapter XIV, p. 178.

[9]Indeed, I do not agree with it. See Priest (1998), esp. 1.5–1.10.

[10]Chapter X, pp. 145–6.

Contradictory objects will loom large in our discussion of the second half of the book. So let us leave further discussion of this example till then.

2. Aristotle's battery of small arguments are easier to understand than the long argument. And concerning these, Łukasiewicz makes two points worth noting (Chap. XII).

First, several of these arguments have a conclusion to the effect that it is not the case that *all* contradictions are true. Łukasiewicz notes, correctly, that this conclusion is beside the point: what needs to be established is that it is not the case that *some* contradictions are true. In these arguments, then, Aristotle has made an illicit slide from *some* to *all*.

Secondly, several of these arguments deduce supposedly unacceptable consequences of violating the PNC, and then apply *modus tollens*. Such arguments fail, says Łukasiewicz, since *modus tollens* presupposes the PNC. Suppose that A entails B and that ¬B. Then without the PNC one cannot infer ¬A, for we might simply have B and ¬B.

Łukasiewicz' point is moot. If entailment is defined (as usual) in terms of the preservation of truth forward, it is correct. However, if it is defined as the preservation of truth forward *plus* the preservation of falsity backward, *modus tollens* in perfectly acceptable. However, I note the point only to show that Łukasiewicz holds that standard inferences concerning negation may fail without the PNC.

3. Łukasiewicz notes and defends Aristotle's claim in the *Analytics* that syllogistic validity does not presuppose the PNC (Chap. XV).[11] Indeed, in Chap. XVI, we find a remarkable thought experiment aimed to show that most reasoning does not require the PNC. Łukasiewicz asks us to consider a society where people take every negated sentence to be true (and so some things are both true and false). Such people would not care about negation at all, but could still reason by induction and by syllogism. Thus, a doctor can recognise the symptoms of diphtheria, remember that a certain drug has been successful at curing such symptoms, and so prescribe it. None of this concerns negation. He concludes (p. 191):

> The example shows that beings that do not recognize the principle of contradiction, that ascertain matters of experimental fact, are able to reason inductively and deductively and are able to act effectively on the basis of such conclusions. If, however, these thought processes are possible in *one* case, then they must be possible in *all* cases. If, by the way, the intellectual organization of our fictional beings would not be different from the human one, then they would be in a position to develop the same sciences as the ones developed by man. From that society, a second Galileo would emerge who would calculate the paths of balls rolling along tilted chutes and who would postulate the laws of free falling objects based on the foundation of these facts; there would be a second Newton, who would synthesize the discoveries of Galileo, Kepler and Huygens into one unified account by determining the highest principle of mechanics. There would be a second Lavoisier…

[11] *An. Post.* 77a10-23. Aristotle's claim bears a small—and opaque—qualification. Łukasiewicz argues that this does not seriously limit the categorical claim.

and so on. Indeed, we would have scientific business as normal.

The conclusion is somewhat surprising. It would seem obvious to suppose that sometimes scientific testing involves the refutation of hypotheses. This requires *modus tollens*, and Łukasiewicz's himself has already said that this does not work without the PNC. What is happening here? I take it that the model of science that Łukasiewicz is working with is an "inductivist" one.[12] Science starts by making observations. Thus, given that the observed As are $a_1,\ldots a_n$, and that $Ba_1,\ldots$, Ba are observed, we may infer that all As are Bs. We can then infer further things my deduction. Thus, if we have also established that all Bs are Cs, we can infer that all As are Cs. Few would now subscribe to this view of science. But it must be remembered that when Łukasiewicz was writing, Popper's *Logic of Scientific Discovery* was still 24 years into the future. However, again, I note Łukasiewicz' view, not for its correctness, but for its relevance to matters to emerge.

4. Łukasiewicz's closes the chapter, and so the first part of his book, with a summary worth noting. Amongst its points are the following[13]:

> b) The principle of contradiction is not a *final* law but it demands proof. c) Aristotle has not provided proof because his arguments were insufficient. Thus, as long as no one else delivers a proof, the principle of contradiction remains an *unjustified* principle in which we have blind faith. f) There are cases in which the principle is certainly *false*, namely with respect to contradictory objects.

Łukasiewicz' Proof

To the second part of the book, then, which contains Łukasiewicz' positive considerations on the PNC.[14] This begins in Chap. XVII, as follows (p. 194):

> I have completed the primarily critical part of the investigations. The more the results turned out to be negative, the stronger the need grew to add a positive part. Despite all this, nobody seriously doubts the principle of contradiction.

His ambivalence concerning the PNC is already clear to the not-so-discerning eye. How could Łukasiewicz fail to doubt the principle, given that he concludes the previous chapter by saying that there are cases where it is 'certainly false'!

At any rate, this chapter contains Łukasiewicz' proof promised in the Introduction to the book. Right at the start of the book, he defines an object as any *something*[15]:

---

[12]See Chalmers (2013), esp. Chap. 1.

[13]Chap. XVI, p. 192.

[14]In the first half of the book Łukasiewicz finds three versions of the PNC in Aristotle. Logical: A and ¬A cannot be true together. Metaphysical: an object cannot both have and not have a property. Psychological: no one can believe A and ¬A. The third, he argues, is just factually false. The first two are, however, equivalent, and provide the subject of the following discussion.

[15]Chap. I, p. 89.

> *By object I mean any something whatever that is "something" and not "nothing"...*

In the present chapter, we then find[16]:

> There is only *one* way… [to prove the PNC]: it has to be assumed that contradictory objects are no objects at all, that they are not something but *nothing*. Anything, then, that is an object and therefore is something and not nothing, does not contain contradictory properties… And here we have the proof of the principle of contradiction, *the only strict and formal proof* that, in my opinion, does exist.

I'm sure that the reader will not be impressed. Neither was Łukasiewicz. He says (p. 201) 'I doubt that this proof will deceive anyone', and then goes on to explain why (p. 202):

> According to the first definition we will call an "object" everything that is something and not nothing, i.e., things, persons, phenomena, events, relations, the entire external world and everything that takes place within ourselves. Also all scientific concepts and theories are objects. According to the second definition, we will call everything an "object" which does not contain a contradiction. The question arises: *are objects in the first sense also objects in the second sense*?… This is the real problem and we have been looking for its solution from the beginning.

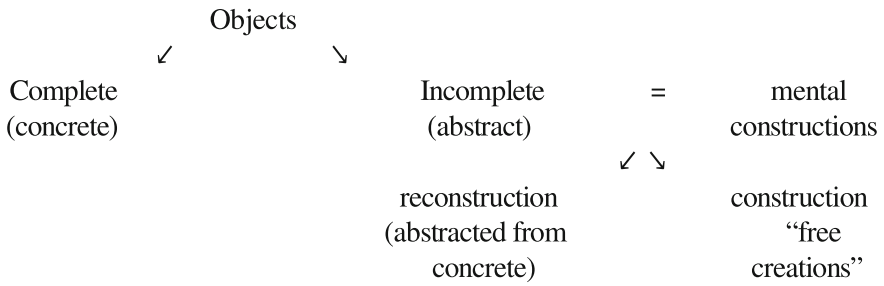
And it is the one to which Łukasiewicz turns in the subsequent chapters.

Impossible Objects

He begins his discussion in Chap. XVIII with a taxonomy of objects. First, there is a division between *complete objects*—i.e., objects, x, such that for any property, P, either Px or ¬Px—and *incomplete objects*. (So far, no commitment to whether one can have, for either kind, an object such that Px and ¬Px.) Concrete objects are complete. Abstract objects are incomplete. Abstract/incomplete objects do not exist in reality, but are 'merely products of the human mind' (p. 205).

Incomplete objects themselves are of two kinds. *Reconstruction objects* are those obtained by abstraction from concrete objects. Thus, *the tree as such*, is obtained from trees by abstracting away all properties which some trees have and some trees lack, leaving the rest. Thus, *the tree as such* has a wooden trunk and bears leaves. It is neither deciduous nor not deciduous; it is neither evergreen nor not evergreen. *Construction objects*, on the other hand, are the 'objects of a priori concepts, which are primarily the concern of mathematics and logic' (p. 205), such as numbers and geometric figures. These are free creations of the human mind (p. 205). Hence we have the following taxonomy:

[16]Chap. XVII, p. 201.



Obviously one may take issue with Łukasiewicz' taxonomy. Thus, for example, platonists about abstract objects will standardly hold that numbers are neither incomplete nor free mental creations. However, this is not relevant here. We are simply tracking Łukasiewicz' thought at this point.

The consistency of concrete objects is taken up in the next chapter, which also comments on reconstruction objects. The rest of the present chapter is taken up with the issue of the consistency of construction objects. The consistency of such objects, it seems, is easily established. Since such objects are free creations of the human mind[17]:

> we have unlimited freedom in their construction… it depends only on us whether these objects turn out contradictory or not contradictory: but because we do believe in the principle of contradiction, we construct them in such a way that they would not be contradictory. Accordingly, we may assert at least about construction objects with certainty that none of them can simultaneously contain and not contain a property.

This is rather surprising. What one might have expected him to say is that since they are entirely free creations, of course we can construct contradictory objects if we wish. But in any case, Łukasiewicz backtracks immediately (p. 206):

> even in the domain of such objects contradictions nevertheless do occur. It is enough to mention "the greatest prime number" or the "square constructed with ruler and compass that has the same area as a circle of radius 1".

The influence of Meinong on Łukasiewicz' is obvious.[18] He is endorsing the Meinongian principle now often called the Characterisation Principle (CP): the thing which is so and so, is indeed so and so. Thus, an object characterised by an inconsistent condition is indeed inconsistent. The CP is a vexed principle, and no one can endorse it without triviality.[19] But such vexations are not on Łukasiewicz' horizon; and given that, one might think, it settles the matter once and for all.

[17]Chap. XVII, pp. 205–6.

[18]Łukasiewicz attended lectures by Meinong in Graz in 1908–1909 (see Simons 1989).

[19]See Priest (2005), esp. the Preface.

But Łukasiewicz' takes away with one hand what he has given with the other (p. 206):

> To this one can reply that these contradictory objects, which obviously are *not* objects, have found their place among other constructions only erroneously and by accident, and because our imperfect intellect is unable to grasp the entire manifold of properties and relations in a single moment and cannot in all cases detect a contradiction right away. But we have immediately removed such objects from science as soon as it became apparent that the objects mentioned were contradictory, and nowadays we already know that the squaring of a circle is impossible and that the greatest prime does not exist.

The remark is doubly puzzling. In the case of the two examples given, their inconsistency is not, indeed, immediately apparent. But what of *the round square* and its like? If one endorses the CP, then this is round and square: its inconsistency is patent. Perhaps, then, it was never constructed mentally? But it seems that it *must* have been constructed if we can think of *it*. Worse, even if it be the case that examples of the kind given are removed when we discover that they are contradictory, it remains the case that they *were* constructed in the first place, and were contradictory. Something done by accident, such as an insult, has still been done.

Łukasiewicz' seems oblivious to this, but raises another problem of his own making (p. 206):

> But the doubt remains: If we are unable to recognize contradictions right away, then how do we know that constructions, which are held to be not contradictory, do not contain a contradiction? Perhaps we have not discovered it until now. This doubt can be expressed in the form of a charge of principle: *Where is the guarantee that non-contradictory objects exist at all?*

He points out, quite correctly, that things constructed can have properties that go beyond those explicit in their construction. Thus, that it is contradictory is no part of the explicit characterisation of the greatest prime number. So how does one know of any construction that it is consistent? He continues (p. 207):

> And once again, someone might object at this point that if every construction were to contain a contradiction, we then would, more frequently than up to now, encounter contradictory objects. In the meantime, one should consider the examples just mentioned merely *exceptions*. They are simply the leftovers in the workshop of science, impurities on the surface of grey molten iron.

Łukasiewicz' remarks are exceptionally puzzling. He moots a doubt to the effect that all construction objects are contradictory. This is irrelevant in the context: it is a slide from *some* to *all*, of a kind for which Łukasiewicz himself has rightly castigated Aristotle. More: even if the worry is real, it fails to address the question on the table: are *some* objects contradictory? Indeed, the passage seems to concede that they are. If there are exceptions, then there are *exceptions*.

Paradox

Łukasiewicz fails to note any of this, but takes issue, instead, with the suggestion that the constructions are mere intellectual flotsam (p. 207):

> However, that this nevertheless is not the case, and that the purity of the metal itself is strongly under suspicion, that is testified by the newest investigations of the foundations of mathematics.

Contradictory objects are not metallic froth: they are at the very core of mathematics, in the form of the theory of sets.

Łukasiewicz shrewdly observes (p. 209) that solving the traditional paradoxes of the infinite by enforcing the condition that two sets are the same size iff they can be put in one to one correspondence, has merely succeeded in shifting paradoxes from the infinite to the absolute infinite—in the form of paradoxes such as Russell's and Burali-Forti.[20]

Łukasiewicz says that he *suspects* that a solution to Russell's paradox, compatible with the Principle of Excluded Middle, can be found (p. 212),[21] but he clearly has nothing to offer here. He then continues:

> I want to return to the problem to which this chapter is dedicated. We did ask whether construction objects, that is, the a priori concepts of mathematics and logic, are objects in the second sense of the word, which is to say, whether they contain contradictory properties or not. The presented examples show that we cannot answer the question unequivocally. In fact, we do encounter strange contradictions in these objects and can never know with certainty whether apparently contradiction free objects really are such.

His conclusion to the chapter, then, is one of agnosticism (p. 213):

> *In fact, we do encounter strange contradictions in [construction] objects and never can know with certainty whether apparently contradiction free objects really are such.*

But it is hard to avoid the conclusion that Łukasiewicz is loth to accept the force of his own arguments. Twice in the chapter he has had counter examples to the PNC at his finger tips, and twice he has backed away from them. In the first case, the examples concerned contradictory Meinongian objects. These counter-examples were rejected by what can only be described as disappointingly sloppy thinking from a mind as acute as Łukasiewicz'. In the second case, these concerned the paradoxical objects of set theory. The best he can do concerning these is to express the hope that there is something wrong with the arguments concerned. But in a context where the PNC is at issue, and so cannot be assumed, this would appear to betray another failure of rationality. The wise person, as Hume noted,[22] apportions

[20] See Priest (1995), p. 126 of 2nd ed.

[21] He notes that the argument for the Russell paradox uses the Principle of Excluded Middle: the Russell set is either a member of itself or it is not. Since construction objects are incomplete, perhaps he had doubts about this.

[22] *Enquiry Concerning Human Understanding*, Part X, Sect. 1.

their beliefs according to the evidence. The evidence available to Łukasiewicz in this case—less than conclusive though it may be—is that these objects really are contradictory.

Motion

The next chapter, XIX, turns from the consistency of abstract objects to that of concrete objects. Łukasiewicz starts by saying (p. 214) that though reconstruction objects are mental constructions, because they are abstractions from concrete reality, they can be contradictory only if it is so. Could it be? Łukasiewicz says (p. 214–5):

> There does not seem to be anything easier than the answer to the question… If there is anything that cannot be doubted, then it is *this* fact: real existing phenomena, things and their qualities, do not contain contradictory properties. If I am now sitting at my desk and write, then it cannot be true at the same time that I am not sitting and writing… [There is then a series of such homely examples.] In fact, such and similar considerations taken from daily life are the strongest arguments for the principle of contradiction.

The strongest argument, then, is one by induction. But of course, even if such an induction is valid, it establishes the principle only for medium-sized dry goods; and Łukasiewicz is well aware, as we have seen, that the PNC might have only limited validity. Indeed, he indicates this almost immediately—in a grudging sort of way (p. 215):

> Things are completely different, if someone is not satisfied with the merely superficial considerations of the appearances and engages in a more subtle analysis. Whoever does this, moves away from "healthy common sense" and has only to blame himself if he gets caught in contradictions.

Łukasiewicz' lead-off batter in such less superficial considerations is Zeno of Elea, whose arguments appeared to establish the contradictory nature of motion. It was not only some of the presocratics who were persuaded. So was Hegel, whom Łukasiewicz quotes as saying[23]:

> [One] has to concede to the old dialecticians the contradictions that they demonstrate in motion, but from this it does not follow that motion does not exist, but rather than motion is the being-present of contradiction itself…

He then adds—somewhat in tension with his discussion of paradoxical objects in the previous chapter:

> Thus, it appears that if it is possible to cast doubt on the principle of contradiction at all, then it will be in the area of concrete objects, that is, in the area of facts of experience.

[23]p. 216. Łukasiewicz' text then refers back to Chap. V, where he has quoted a passage from Hegel's *Logic* expressing the point even more explicitly.

What, then, is one to say of apparent contradictions concerning motion? Says Łukasiewicz, they may be resolved by drawing an appropriate distinction: time. 'Concrete objects may contain contradictory properties, but not "simultaneously", that is, not at one and the same time' (p. 216). Perhaps. But Łukasiewicz is troubled by Zeno's arrow paradox (pp. 217–8):

> Let us imagine a transverse section cut across the entire world of phenomena, performed at some arbitrary point of time. In this transverse section, on its immobilized surface, there would be no change and no time. The arrow would have to freeze motionless at some location. But how do we know that it would have to be at only *one* location? As long as it was moving, it continuously changed its location in space and, consequently, it was present at many locations even within each smallest of time moments. Why, then, could it not also be at least at *two* different places in the not extended time-point of the transverse section?

Łukasiewicz goes on to answer his own question (p. 218):

> There can be no answers to these questions. One cannot gain anything here with a priori considerations because one would have to already rely on the principle of contradiction, which is what we are trying to ground. Experience, too, is silent on the matter, since a not extended time-moment is not an object of experience.

But he misses the obvious here. Zeno does not just raise the possibility that the arrow is instantaneously in a contradictory state. He gives an argument for it: if it made no advance at each instant of its journey, it could make no advance at all.[24] And just as Łukasiewicz would have to find an error in the argument for Russell's paradox, he would have to find an error in Zeno's argument. It does not seem to occur to him that he needs to do this. At any rate, Łukasiewicz draws the same agnostic conclusion he draws in the previous chapter: one just cannot know whether reality is inconsistent.

But Łukasiewicz, torn, seems to feel that he cannot leave it at that. He then says —contradicting what he said earlier (after his quotation from Hegel) about the realm of the concrete providing the most likely counter-examples to the PNC (p. 219):

> But despite all this, the case of the principle of contradiction is stronger in the domain of real objects than in the sphere off mental constructions. There we encountered *factual* contradictions, whose solution is not at all easy, and here, on the other hand, the existence of the contradiction is merely *possible*.

Moreover, he says, if at some time we seemed forced to conclude that an object were in two places at the same time, we could just conclude that being in two places at the same time is not a contradiction. After all, if p_1 and p_2 are distinct places, 'x is at p_1 and x is at p_2' is not a literal contradiction. For that, we need the extra premise that if x is at p_1, it is not at p_2.[25] He concludes that we might treat the principle of

[24] See Priest (1987), 12.2.

[25] Łukasiewicz makes the point only concerning the experience of an object being in two places at the same time. But he might equally have applied the thought to instantaneous situations, thus attempting a solution to the Arrow Paradox.

non-contradiction in the same way that we treat the principle of causation: everything has a cause. If we ever come across an apparent counter-example to this, we can always assume that there is a cause; we have just not found it yet. The two principles, though, are evidently not the same. Were we to find ourselves in the situation of having to suppose that an object is in two places at the same time, we do not have to suppose that there is something going on which we have not yet discovered: we just have to take it, according to Łukasiewicz, that being at two different places at the same time is not a contradiction.

In any case, this supposed resolution of the supposed contradiction is somewhat lame. Of course, an *extended* object can be in two places at the same time: my left and right hands are at different points in space. But the very notion of a point is one which can have no such extension. Its being in more than one place is ruled out by very definition. And any point on the arrow—which moves in tandem with the rest of the arrow—is such a point.[26]

The chapter closes with the following paragraph (p. 221):

> The final result of the last two chapters has in principle been negative: it is impossible to show in an indubitable manner that contradictory objects exist. Long centuries of scientific work separate us from the moment during the origins of philosophy when Aristotle sought to prove the existence of at least one contradiction-free being. Today we are older, and thus more modest.

The statement is telling. We cannot prove the existence of a non-contradictory object. Perhaps so. But that was not what was at issue. What was at issue is whether we could prove that all are contradict-free. Like Aristotle sliding from *some* to *all*, Łukasiewicz, sensing that he cannot prove what he has sought, slides to a different claim, without remarking on the fact.

A "Practical-Ethical" Principle

In the last substantive chapter of the discussion, Łukasiewicz addresses two questions concerning the PNC[27]:

(a) Why is it that we believe in a principle whose truth cannot be demonstrated?

(b) Why do we attribute a value to it that exceeds even the value attributed to statements that are true with certainty?

The answer to (a) is simply the magisterial authority of Aristotle (p. 224):

> It is a mark of the genius of the Aristotelian spirit, that it was able to convince all humanity of two things: First, that the principle of contradiction is true *even though there is no proof*

[26]One might, I suppose, suggest that the length which is the arrow, is not made up of points. So much the worse for contemporary science. And supposing spatial distance to be quantized drives us into the arms of another of Zeno's paradoxs anyway: the Stadium.

[27]Chap. XX, p. 230.

> *for it*; and further that the principle of contradiction *does not require a proof at all*. Has there ever been anything comparable in the history of science?

That seems about right. The problematic nature of Aristotle's arguments is patent to anyone who considers them with an open mind; and there has been no substantial defence of the PNC since. What was there, then, but Aristotle's name?

The answer to (b) is more substantial, and it is here that Łukasiewicz finally lays his own cards about the PNC on the table. He says (p. 226):

> … this is the place to present the final and probably most important idea of this treatise. It seems to me that so far nobody has brought this thought into clear awareness, even though Aristotle was perhaps closest to it: *the value of the principle of contradiction is not of a logical but of a practical-ethical nature: this practical-ethical value, however, is so great that the lack of logical value does not count in comparison*.

Łukasiewicz illustrates what he means. Suppose that I am accused of a crime. I can show that I was not there. I have an alibi, a witness who will vouch that I was somewhere else, and so on. But another witness comes forward who swears that I was there, and that he saw me commit the crime. Given the principle of non-contradiction, both witnesses cannot be correct, and the judge has to decide who is the more reliable witness, where the balance of evidence lies, etc. But without the principle, the judge may just decide that I was there, even if I wasn't, and find me guilty.[28] The example (p. 228):

> show[s] what the practical and ethical significance of the principle of contradiction consists in. The principle is the only weapon against mistakes and lies. If contradictory statements were to be reconcilable with each other, if affirmation were not to nullify denial, but if the one were to meaningfully co-exist next to the other, then we would have no means at our disposal to discredit falsity and unmask lies.

This is a somewhat amazing argument for Łukasiewicz to use. As we have seen, both he and (according to him) Aristotle have noted that the PNC might hold only in a limited realm. The fact that we may apply the principle in this case says nothing, therefore, about whether it applies to mental constructions and instantaneous states, for example. Łukasiewicz even comments on the fact a page later, where he says (p. 229) that there is no danger to the empirical sciences 'if some contradiction in the a priori sciences, for example, Russell's contradiction cannot be resolved'.

Indeed, it is not even true that one needs the *impossibility* of contradiction in this case. The *improbability* will do. As Hume noted in his discussion of miracles,[29] we may well justifiably refuse to accept that something has happened if it is most unlikely—especially if there is a better explanation of the situation (in this case, that someone has lied). After all, even if everything is entirely consistent, given the laws

[28] What my evidence must do is cause the judge to reject the claim that I was at the scene of the murder. However, the distinction between asserting a negated sentence and denial—the linguistic correlate of rejection—is not on Łukasiewicz' radar. (For more on the distinction, see Priest 1987, 7.3, 2006, 6.2.).

[29] *An Enquiry Concerning Human Understanding*, Sect. X.

of quantum mechanics, there is some probability that my atoms might have collectively disappeared and rematerialised at the scene of the crime long enough for me to pull the trigger. Pity the poor barrister for the prosecution who tried to use this as an argument!

Łukasiewicz then generalises his point to a more grandiose scenario: the development of science in Ancient Greece. At that time the new empirical sciences were coming into being, but many of the sophists denied the PNC. Had Aristotle not defended it, the nascent science would have been still-born. Commenting on Philip of Macedonia's defeat of the Thebans and Athenians, he explains, in the following colourful passage, that Aristotle must have (p. 230–2):

> felt heart break over the battle of Chaeronea. In this politically most difficult moment, which made practical life so distasteful to Aristotle and encouraged him to treasure theoretical life above all, the sophists introduced the elements of intellectual and moral laxness. This constituted a hundred times greater defeat than the might of Macedonia. It not only destroyed the foundations of social being but also the foundations of the individual; it destroyed the principles of the understanding. Aristotle saw the future of his homeland in the cultural work that was to remain the only free area of activity for the Greeks for centuries to come. And he took on an inexhaustible part of this labour by establishing the considerable foundations of a *scientific* culture and, by combining his research and that of others, created a series of new and systematic branches of knowledge. The sophists, in particular, had proven themselves as the enemies of such goal-oriented, creative, and systematic labour… The sophistries and paradoxes of these clever speakers were known all over Greece. Perhaps there wasn't anyone who took these strangely distorted thoughts too seriously, but they nevertheless ridiculed science in the public eye and instigated chaos in the minds of men. These sophists denied the principle of contradiction. Even though their charges were vacuous, the positive proof of the principle was not possible and… Aristotle was himself aware of the weakness of his arguments. Thus, there was nothing else left than to declare the principle of contradiction a *dogma*, and to set authoritarian limits to any destructive works. Only in that way was the Stagirite able to forge armours against sophistries and errors and the clear the path for positive work.

Whether or not this is true, it is a bizarre statement for Łukasiewicz to make, precisely because, as we have already noted, he has argued a few chapters before that scientific development would continue in exactly the same way, even without the PNC. Aristotle did not need the dogma!

Conclusion: *Tu Quoque*

Let us draw the threads of our discussion together. Łukasiewicz clearly accepts the PNC; indeed, he declares that no one doubts it, and asserts confidently at the start of the inquiry that it can be proved. But as the investigation proceeds, he concedes that the proof he gives is of little value. Objects, by definition, may be consistent, but the real question is then whether there are some things that are not objects. He discusses contradictory Meinongian objects, paradoxes of set theory, and the views of Zeno and Hegel on motion. His official conclusion is that it is impossible to show these situations do not generate contradiction. But in many places he seems forced to

concede that there are arguments for contradiction that he does not know how to answer—sometimes he does not even try. Indeed, at one point earlier in the investigation, he actually lets slip the admission that the PNC is indeed false. Finally, he resorts to a pragmatic justification of the PNC which is not only rather feeble, but fails given his own prior considerations.

In the first half of the book, commenting on Aristotle's arguments for the PNC, he says[30]:

> It seemed as if the Stagirite, filled with trust in his own strength and certain of victory, entered head-on into a fight. Like arrows from a quiver, he takes out one proof after the other; but he does not notice that the arrows have no effect. He has exhausted the arguments and none of them has demonstrated the principle that was so dear to him. So he defends with what is left of his strength and faith the last position: that of only one contradictory free being and only one contradiction free truth.
>
> Or perhaps it has been different? Perhaps his pride and self-assuredness were just pretended? Sometimes one wants to assume that Aristotle, sensing the practical and ethical importance of the principle of contradiction with his acute and deep understanding, *deliberately* formulated it as an untouchable dogma in order to replace the lack of factual arguments with his *sic volo sic iubeo*. But in the depth of his soul, he himself was not sure about this matter. He hid himself behind this thought; the debate, however, made him grow passionate. And in this way, he let slip a moan of despair against his will… It will be shown soon that he had in fact enough reasons to doubt the universal significance of the principle of contradiction, but apparently did not have enough courage to admit this outright.

Though Łukasiewicz is speaking of Aristotle, he might just as well have been speaking of himself.[31]

References

Chalmers, A. (2013). *What is this thing called science?* (4th ed.). Brisbane: University of Queensland Press.

Heine, H. (2013). Jan Łukasiewicz and the principle of contradiction (Ph.D. Thesis, University of Melbourne, 2013).

Łukasiewicz, J. (1910a). O Zasadzie Sprzeczności u Arystotelesa. Kraków: Akademia Umiejętności (2nd ed.). In J. Woleński (Ed.), Warsaw: PWN, 1987. Translated as: Über den Satz vom Widerspruch bei Aristoteles. Hildesheim: Olms, 1993. Del Principio di Contradizzione in Aristotele. Macerata: Quodlibet, 2003.

Łukasiewicz, J. (1910b). Über den Satz von Widerspruch bei Aristoteles. Bulletin Internationale de l'Académie des Sciences de Cracovie, Classe de Philosophie, 15–38. Translated as: On the Principle of Contradiction in Aristotle. Review of Metaphysics 24: 485–509 (1970/71). Aristotle on the Law of Contradiction. Articles on Aristotle 3. Metaphysics, eds. J. Barnes, M. Schofield, and R. Sorabji, 50–62. London: Duckworth, 1979.

Priest, G. (1987). *In contradiction*. The Hague: Martinus Nijhoff (2nd ed.). Oxford: Oxford University Press, 2006.

[30]Chapter XIII, pp. 170–1.

[31]A draft of this paper was read to the Melbourne Logic Group, July 2014. Thanks go to the members of the group for their helpful discussion.

Priest, G. (1995). *Beyond the limits of thought*. Cambridge: Cambridge University Press (2nd. ed., Oxford: Oxford University Press, 2002).
Priest, G. (1998). To Be and not to be—that is the answer: Aristotle on the law of non-contradiction. *Philosophiegeschichte und Logische Analyse* 1: 91–130. (Reprinted as ch. 1 of Priest 2006).
Priest, G. (2005). *Towards non-being*. Oxford: Oxford University Press.
Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Oxford University Press.
Simons, P. (1989). *Łukasiewicz, Meinong and many-valued logic*. In K. Szaniawski (Ed.), The Vienna circle and the Lvov-Warsaw School (pp. 249–91). Dortrecht: Kluwer.

Author Biography

Graham Priest is Distinguished Professor of Philosophy at the Graduate Center, City University of New York, and Boyce Gibson Professor Emeritus at the University of Melbourne. He is known for his work on non-classical logic, particularly in connection with dialetheism, on the history of philosophy, and on Buddhist philosophy. He has published articles in nearly every major philosophy and logic journal. His books include: *In Contradiction: A Study of the Transconsistent*, Martinus Nijhoff 1987 (2nd edition: Oxford University Press 2006); *Beyond the Limits of Thought*, Cambridge University Press 1995 (2nd edition: Oxford University Press 2002); *Towards Non-Being: the Semantics and Metaphysics of Intentionality*, Oxford University Press 2005; *Doubt Truth to be a Liar*, Oxford University Press 2006; *Introduction to Non-Classical: Logic From If to Is*, Cambridge University Press 2008; *One*, Oxford University Press 2014. Further details can be found at: grahampriest.net.

Why Did Weyl Think that Dedekind's Norm of Belief in Mathematics is Perverse?

Iulian D. Toader

This paper discusses an intriguing, though rather overlooked case of normative disagreement in the history of philosophy of mathematics: Weyl's criticism of Dedekind's famous principle that "In science, what is provable ought not to be believed without proof." The analysis of this principle will identify two main assumptions: that it is rational for one to want to improve one's believing, or as I will prefer to put it, one's doxastic performance in mathematics, and that minimizing the extent of the probative contribution of immediate insight or intuition actually improves this performance. After noting Peirce's reasons for rejecting the former, I argue that Weyl rejected only the latter, for he thought that immediate insight or intuition was a cognitive asset, not a liability, and that minimizing the extent of its probative contribution worsened, rather than improved, doxastic performance. This criticism, as I see it, challenges not only a logicist norm of belief in mathematics, but also a realist view about whether there is a fact of the matter as to what norms of belief are correct.

As is well known, in his 1918 book on the foundations of real analysis, *Das Kontinuum*, Weyl attempted to show that it is possible to justify the definition of real numbers as sets of rational numbers—a definition inspired by Dedekind—only if impredicative definitions are verboten. This restriction, as Weyl readily acknowledged, entails that some theorems of classical analysis, like the least upper bound theorem for sets of real numbers, cannot be proved. Consequently, he had to give up the Dedekind completeness of the real numbers. But Weyl was able to predicatively prove the least upper bound theorem for sequences of real numbers. He further offered predicative proofs of other important results, for example, that a function which is continuous on the unit interval has a maximum and a minimum in that interval. And although he did not provide a proof, he also correctly noted that this result is enough to prove the fundamental theorem of algebra.

I.D. Toader (✉)
University of Bucharest, Bucharest, Romania
e-mail: itoad71@gmail.com

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9_19

Less than well known is that Weyl also recorded, in the same book, his rejection of Dedekind's ideal of probative completeness, that is the imperative to provide a logical foundation to all provable propositions.[1] He did so in the following footnote: "In the Preface to the first edition of Dedekind's famous *Was sind und was sollen die Zahlen*?, we read that 'In science, what is provable ought not to be believed without proof.' This remark is certainly characteristic of the way most mathematicians think. Nevertheless, it is a perverse principle. As if such a mediated concatenation of grounds as what we call a 'proof,' can awaken any 'belief' without our assuring ourselves, through immediate insight, of the correctness of each individual step! This (and not the proof) remains throughout the ultimate source of knowledge; it is the 'experience of truth'." (Weyl 1918, 11. Eng. tr., 119.) Despite being often quoted in the literature, this footnote has not really been given the philosophical attention it deserves. Some authors refer to it in order to emphasize the general importance of immediate insight or intuition in Weyl's overall thinking (e.g., Bell 2000). Others think that the footnote reveals, more specifically, his early adoption of a phenomenological approach to science and his criticism of the formal axiomatic approach (e.g., Ryckman 2005). But I think that Weyl's argument, as well as its philosophical significance, are still in need of clarification.

Let's start by looking more closely at Dedekind's principle. This was initially intended as a motto to his 1888 book,[2] which suggests at the very least that it was thought to have a central significance for his logicist program in the foundations of mathematics. The principle expresses, I think, a certain norm of belief in mathematics, but before we clarify what it means and how it may be understood, it is worth noting that Dedekind did not deny that one can believe a provable proposition without proof, as one surely does. He denied only that believing it without proof may be considered epistemically adequate, in a sense to be presently clarified.

Dedekind's principle assumes that one can separate what is provable, from what is not, and thus that one can provide a description of the epistemic conditions that would allow one to realize that a proposition is unprovable, i.e., that it is, as he put it, "a pure law of thought."[3] Aristotle, let us recall, had suggested that a certain type of education might help one justify this assumption, at least with respect to some propositions: "It is impossible for anything at the same time to be and not to be. […] This is the most indisputable of all principles. Some indeed demand that even this shall be demonstrated, but this they do through want of education, for not to know of what things one should demand demonstration, and of what one should not, argues want of education."[4] Whereas Aristotle thought that education was needed to guard against those who want to prove too much, Dedekind considered

[1]The expression "probative completeness" has been introduced in Detlefsen (2010) to characterize an epistemic ideal shared, before Dedekind, by Bolzano. See also Detlefsen (2011).

[2]See Dedekind's letter to H. Weber from November 19, 1878; in Dedekind (1932), 486.

[3]It is, of course, difficult to give a compelling delineation of the class of unprovable propositions. As Frege duly noted, Dedekind himself offered no "inventory of the logical or other laws taken by him as basic." Cf. Frege (1893, viii).

[4]Cf. *Metaphysics*, Book IV, Chap. 4.

that the nineteenth century methods of teaching elementary mathematics encouraged one to leave too much without proof: "So from the time of birth we are continually and in increasing measure led to relate things to things […] this exercise goes on continually, though without definite purpose, in our earliest years; the accompanying formation of judgements and chains of reasoning leads us to a store of real arithmetical truths to which our first teachers later refer as to something simple, self-evident, and given in inner intuition."[5] Dedekind blamed such teaching methods for encouraging one to consider as simple, self-evident, and immediately presented to the mind, and thus to believe without proof, that which is only the result of judgement and reasoning. On his view, considering arithmetical truths in this way conceals the "chains of reasoning" behind them. Scientific belief requires, on the contrary, that these chains be fully revealed. This was, according to him, the task of mathematical proof. This task, once accomplished, would allow one to realize that "the concept of number [… is…] an immediate outflow (*einen unmittelbaren Ausfluss*) from the pure laws of thought […i.e., of] the ability of the mind to relate things to things, to let a thing correspond to a thing, or to represent a thing by a thing, an ability without which no thinking is possible. […] Upon this unique and therefore absolutely indispensable foundation […] must, in my judgement, the whole science of numbers be established." (ibid.) Dedekind seems to have further believed that this ability, i.e., the very ability that, when exercised long enough, leads to arithmetical truths, is something that needs to be given immediately to the mind, in inner intuition. But he does not seem to have been interested in offering a more definite view about the epistemic conditions that characterize this ability.

Let's call a proof that provides a logical foundation to a provable proposition, in Dedekind's sense, a d-proof. Thus, for example, a d-proof of proposition $m+n=n+m$, which is theorem 140 in his system of arithmetic, requires not only arithmetical complete induction, i.e., the inference from n to $n+1$, which is theorem 80 in the system, but also generalized complete induction, which is theorem 59. A proof of $m+n=n+m$ that proceeded just by arithmetical complete induction would presumably be inferentially deficient, for it would conceal the chain of reasoning behind theorem 80. As Dedekind noted, theorem 80 "results immediately" (*ergiebt sich unmittelbar*) from theorem 59. But even if theorem 59 were deployed in the proof, the proof would still be deficient, in case the chain of reasoning behind theorem 59, up to the pure laws of thought, remained concealed. Thus, Dedekind's principle requires that one eliminate such inferential deficiencies: for any provable proposition p, one ought not to believe that p if a d-proof of p (or a sequence of proofs that is logically equivalent to a d-proof of p) has not been yet given.

Now, if one does believe a provable proposition p without a d-proof of p (or a logically equivalent sequence of proofs), then Dedekind's principle implies that something has gone wrong doxastically. The principle thus seems to make two further assumptions: that it is rational for one to want to do better doxastically, to improve one's doxastic performance, as it were, and that doxastic performance can

[5]Cf. Dedekind (1888). Eng. tr. in Ewald (1996, 792).

actually be improved by updating the proof of p to a d-proof (or to a logically equivalent sequence of proofs).[6] Consequently, one can distinguish two critical approaches to Dedekind's principle: one that rejects his conception of scientific rationality, and another that approves of this conception but rejects his view about what improves doxastic performance.

The criticism that Peirce raised against the type of epistemological project that Dedekind embarked upon illustrates, I think, one approach. Peirce famously wrote: "That the settlement of opinion is the sole end of inquiry is a very important proposition. It sweeps away, at once, various vague and erroneous conceptions of proof." One such conception would be the one later defended by Dedekind, for Peirce went on: "Some people seem to love to argue a point after all the world is fully convinced of it. But no further advance can be made. When doubt ceases, mental action on the subject comes to an end; and, if it did go on, it would be without a purpose." (Peirce 1877, 11) Purposeful mental action (and, thus, proper scientific inquiry) requires the presence of "real and living" (as opposed to mere Cartesian) doubt, which arises only when logical inconsistencies are revealed. Such doubt stimulates the mind to seek the settlement of belief. Once inconsistency is removed, stimulation ceases, belief settles, and inquiry should stop. Thus, in the case of a provable proposition, Peirce's criticism of Dedekind's principle would emphasize that when one's doxastic performance reached a certain level of competence, one should stop seeking to improve that performance. To postpone settlement, which is what Peirce thought was characteristic of Cartesian epistemology, is a "perversity." (Peirce 1878, 39) But to seek improvement after belief has settled, which is what Peirce would say about Dedekind's epistemological project, would be no less irrational.[7]

Weyl's criticism of Dedekind's principle illustrates, I think, the other critical approach identified above. As we have already seen, one assumption behind the principle suggests a certain type of proof revision. That is, for any provable proposition p, if t is a proof of p and t_d is a d-proof of p, Dedekind's principle demands revision to the effect that t is replaced by t_d (or by a sequence of proofs logically equivalent to t_d). What justifies this revision is the fundamental idea that t is inferentially inferior to t_d, where inferential inferiority is due to the kind of deficiencies illustrated in our discussion of theorem 140. Unlike t, t_d reveals the chains of reasoning behind p, up to the pure laws of thought, which according to Dedekind improves doxastic performance by minimizing the extent of the probative contribution of immediate insight or intuition.

So why did Weyl reject Dedekind's principle as perverse, despite the fact that he thought most mathematicians would not? He seems to have agreed with Dedekind that it is rational for one to want to improve one's doxastic performance in mathematics. But on Weyl's view, as extracted from his footnote in *Das Kontinuum*, a

[6]For the view of belief as doxastic performance, see Sosa (2011). My adopting it here without discussion will be remedied in a fuller version of the paper.

[7]For a more general discussion of Peirce's criticism of Cartesian epistemology, see Haack (1982). For a discussion of his view on logicism, see Haack (1993).

mathematical proof improves doxastic performance only if all its individual steps are supported by immediate insight or intuition. However, as we have seen, a d-proof minimizes the extent of the probative contribution of immediate insight or intuition. Therefore, according to Weyl, a d-proof does not improve doxastic performance. So Weyl resisted the type of proof revision demanded by Dedekind's principle because he believed that in order to improve doxastic performance one needed evidentially superior proofs, rather than what Dedekind considered inferentially superior ones. Evidential superiority, according to Weyl, is obtained by maximizing the extent of the probative contribution of immediate insight or intuition. On his view, immediate insight or intuition is a cognitive asset, not a liability; therefore, doxastic performance improves only by maximizing the extent of its probative contribution. For example, in Weyl's predicative system of real analysis, $m + n = n + m$ is proved by arithmetical complete induction, which is then justified in the following way: "the important inferential technique of complete induction is based on the circumstance that one can eventually reach any number whatever by starting from 1 and proceeding from each number to its successor. […] the idea of iteration, i.e., of the sequence of the natural numbers, is an ultimate foundation of mathematical thought—in spite of Dedekind's 'theory of chains' which seeks to give a logical foundation for definition and inference by complete induction without employing our intuition of the natural numbers" (Weyl 1918, 25–48). This suggests that immediate insight or intuition is to be probatively employed as extensively as possible, if one wants to improve one's doxastic performance so that proofs become a genuine source of mathematical knowledge.

It is this argument, I take it, that motivated Weyl's rejection of the logicist norm of belief expressed by Dedekind's principle. But this rejection, if justified, raises the question whether there is a fact of the matter as to what norm of belief in mathematics is correct. In the remainder of the paper, I discuss an argument to the effect that this historical case of normative disagreement is not factual, and then I reply to one objection to it. My all-too-brief discussion is, however, only meant to open the case, rather than closing it.

Consider first the following argument for the nonfactual character of normative disagreement about the validity of the logical law of explosion, i.e., $\{\phi, \neg\phi\} \vDash \psi$: "If I myself believe in the law of explosion but I know that a friend doesn't, I can distinguish between what she should believe about a matter given her logical views and what she 'objectively' should believe about it. Judging what she 'objectively' should believe about validity is straightforward on the view (at least if I ignore the possibility that I myself am misled): it's the same as judging what's valid. Judging what she should believe given her logical views involves some sort of projection, something like an 'off-line decision' of what to believe about the matter in question on the pretense that one's views are in key respects like hers." (Field 2015). This argument seems to say that if the normative disagreement about the validity of explosion is factual, then the kind of projection described here would be impossible or at least misleading. But since projection in this case seems possible and also not misleading, because one does seem to be able to distinguish between different, equally justified, doxastic obligations, it follows that the normative disagreement about the validity of explosion is not factual.

One may defend the nonfactual character of the normative disagreement between Weyl and Dedekind in a similar manner: if the normative disagreement about what improves doxastic performance in mathematics is factual, then one cannot distinguish between different, equally justified, evaluations of doxastic performance in relation to a provable proposition p. But since it seems that one can distinguish between what Dedekind and Weyl should say about a d-proof of p given their own views about what improves doxastic performance, and since their beliefs about the matter in question are equally justified, then it follows that their normative disagreement is not factual.

One immediate objection to this argument is the following: even if it's true that one can distinguish between what Dedekind and Weyl should believe about a d-proof of p given what they think about doxastic performance, one may still doubt that their beliefs about the matter in question are equally justified. To motivate this doubt, consider Dedekind's claim, already mentioned above, that arithmetical complete induction "results immediately" from generalized complete induction. Consider further his contention, also quoted above, that the logicist project of providing d-proofs for all provable propositions of arithmetic reveals the laws of arithmetic as an "immediate outflow" from the pure laws of thought. It seems fair to believe that this contention is based on the idea that a d-proof is such that each of its individual steps is characterized by immediacy, i.e., the type of immediacy that characterizes, according to Dedekind, the derivation of arithmetical complete induction from generalized complete induction. If this is true, then it seems that, contrary to what Weyl thought, a d-proof is not a "mediated concatenation of grounds" and thus not evidentially deficient. Whereas Dedekind's belief that d-proofs improve doxastic performance because they are inferentially superior is justified, Weyl's belief that d-proofs worsen doxastic performance because they are evidentially deficient is not justified.

However, I think that this objection to the nonfactualist reading of our historical case of normative disagreement conflates two different notions of immediacy. A d-proof may admittedly be such that its individual steps are all characterized by immediacy, as Dedekind seems to have intended, while at the same time minimizing the extent of the probative contribution of immediate insight or intuition. But this does not entail that a d-proof eliminates both inferential and evidential deficiencies, in the senses delineated above. For immediacy, as Dedekind seems to have conceived of it, is a property of inference. In contrast, for Weyl, immediacy is a property of insight or intuition. These two notions of immediacy are, so it seems to me, opposite: immediate insight is insight that is not mediated by any inference; immediate inference is inference not mediated by any insight or intuition of, say, the natural numbers. If one distinguishes these two notions, the doubt that Dedekind's and Weyl's beliefs about the matter in question are equally justified is dissipated. So I think that the nonfactualist argument cannot be shot down in this way.

To conclude, the little that has been just said is obviously not enough to claim that the normative disagreement between Weyl and Dedekind *is* not factual, that it is an expression of, say, a merely aesthetic divergence regarding styles of

mathematical proof. But the settlement of belief about this claim must here be postponed (quite perversely, I know) for another stimulating occasion.[8]

References

Bell, J. (2000). Hermann Weyl on intuition and the continuum. *Philosophia Mathematica III, 8*, 259–273.

Dedekind, R. (1888). *Was sind und was sollen die Zahlen?* Vieweg, Braunschweig. Repr. in Dedekind 1932. Eng. tr. in Ewald 1996.

Dedekind, R. (1932). *Gesammelte mathematische Werke* (vol. 3). In R. Fricke et al. (Ed.), Braunschweig: Vieweg & Sohn.

Detlefsen, M. (2010). Rigor, re-proof and Bolzano's critical program. In P.-E. Bour et al. (Ed.), *Construction* (pp. 171–184). UK: King's College Publications.

Detlefsen, M. (2011). Dedekind against intuition: Rigor, scope and the motives of his logicism. In C. Cellucci et al. (Ed.), *Logic and knowledge* (pp. 205–217). UK: Cambridge Scholars Publishing.

Ewald, W. B. (1996). *From Kant to Hilbert*, vol. 2, Oxford University Press.

Field, H. (2015). What is Logical Validity? In Colin Caret and Ole Hjortland (Eds.), *Foundations of Logical Consequence*. Oxford: Oxford University Press (forthcoming).

Frege, G. (1893). *Die Grundgesetze der Arithmetik*, Jena.

Haack, S. (1982). Descartes, Peirce and the cognitive community. *The Monist, 65*, 156–181.

Haack, S. (1993). Peirce and logicism: Notes towards an exposition. *Transactions of the Charles S. Peirce Society, 29*, 33–56.

Peirce, C. S. (1877). "The fixation of belief" in Peirce 1955, 5–21.

Peirce, C. S. (1878). "How to make our ideas clear" in Peirce 1955, 23–41.

Peirce, C. S. (1955). *Philosophical Writings of Peirce*, (Ed.), by J. Buchler, Dover.

Ryckman, T. (2005). *The reign of relativity*. Oxford: Oxford University Press.

Sosa, E. (2011). *Knowing full well*. Princeton: Princeton University Press.

Weyl, H. (1918). *Das Kontinuum*. In *Das Kontinuum und andere Monographien*, Chelsea Publishing Company, 1960. Eng. tr. as *The Continuum*, Thomas Jefferson University Press, 1987.

Author Biography

Iulian D. Toader (PhD, University of Notre Dame, 2011) is a philosopher of science interested in the history and philosophy of mathematics, early analytic philosophy, and naturalistic metaphysics. He published articles in Synthese, HOPOS: The Journal of the International Society for the History of Philosophy of Science, Logique et Analyse, and History and Philosophy of Logic. He edited New Perspectives in the Philosophy of Physics—a special issue of Foundations of Physics, and co-edited the volume Romanian Studies in Philosophy of Science for the Springer series Boston Studies in the Philosophy and History of Science.

---

[8]My understanding of Weyl's philosophy of mathematics benefited from numerous discussions, during my graduate studies, with Paddy Blanchette, Mic Detlefsen, Chris Porter, and Sean Walsh. Since then, I presented this material at several conferences, most recently in Cracow, Rome, and Vienna. Many thanks to my audiences there, and to Hanoch Ben-Yami, Anandi Hattiangadi, and Dirk Schlimm, for their helpful comments and suggestions.

Index

S. Costreie (ed.), *Early Analytic Philosophy – New Perspectives on the Tradition*,
The Western Ontario Series in Philosophy of Science 80,
DOI 10.1007/978-3-319-24214-9