de Gruyter Series in Logic and Its Applications 6

One Hundred Years of Russell's Paradox

Mathematics, Logic, Philosophy

Editor

Godehard Link

Walter de Gruyter
Berlin · New York

Editor
Godehard Link
Philosophie-Department
Ludwig-Maximilians-Universität München
Ludwigstr. 31/I
80539 München
Germany

Series Editors

Wilfrid A. Hodges
School of Mathematical Sciences
Queen Mary and Westfield College
University of London
Mile End Road
London E1 4NS
United Kingdom

Ronald Jensen
Institut für Mathematik
Humboldt-Universität
Unter den Linden 6
10099 Berlin
Germany

Steffen Lempp
Department of Mathematics
University of Wisconsin
480 Lincoln Drive
Madison, WI 53706-1388
USA

Menachem Magidor
Institute of Mathematics
The Hebrew University
Givat Ram
91904 Jerusalem
Israel

Mathematics Subject Classification 2000: 03A05; 00A30, 03-06

♾ Printed on acid-free paper which falls within the guidelines of the ANSI to ensure permanence and durability

Library of Congress − Cataloging-in-Publication Data

One hundred years of Russell's paradox : mathematics, logic, philosophy / edited by Godehard Link.
p. cm. − (De Gruyter series in logic and its applications; 6)
Includes bibliographical references and indexes.
ISBN 3-11-017438-3 (acid-free paper)
1. Paradox. 2. Liar paradox. I. Title: 100 years of Russell's paradox. II. Link, Godehard. III. Series.
BC199.P2O54 2004
165−dc22
2004006962

ISBN 3-11-017438-3
ISSN 1438-1893

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Typesetting using the Authors' TEX files: I. Zimmermann, Freiburg − Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen − Cover design: Rainer Engel, Berlin.

Preface

In June 2001 an international conference was held at the University of Munich, Germany, to commemorate the centenary of the discovery of Russell's paradox. It brought together leading scholars from the following fields:

- Russell Studies
- Mathematical Logic
- Set Theory
- Philosophy of Mathematics

The aims of the event were twofold, historical and systematic. One focus was on Russell's logic and logical philosophy, which were shaped so much by his own and related paradoxes, in particular since new material has become available in recent years from the rich sources of the Russell Archives through the publication of the *Collected Papers of Bertrand Russell*. But the second aim was of equal importance: to present original research in the broad range of foundational studies that draws on both current conceptions and recent technical advances in the above-mentioned fields. It was hoped to contribute this way to the well-established body of mathematical philosophy initiated to a large extent by Russell a hundred years ago.

The conference featured plenary sessions with distinguished invited speakers, section meetings with both invited and contributed papers, and as special events two panel discussions, "Russell in Context" and "The Meaning of Set Theory", as well as two symposia, one on "Propositional Functions" and the other on the "Finite Mathematics of Set Theory". The Russell panel, chaired by A. Irvine, was composed of N. Griffin, P. Hylton, D. Kaplan, and A. Urquhart. Y. Moschovakis chaired the panel on Set Theory, with S. Feferman, H. Friedman, D. C. McCarty, and W. H. Woodin as panelists. The symposiasts of the first symposium, chaired by R. Wahl, were G. Landini and B. Linsky, those of the second S. Lavine and K.-G. Niebergall, chaired by the editor.

The papers collected in this volume represent the main body of research emanating from the proceedings of the conference. All authors delivered a talk at the conference except J. Mycielski, who was unable to attend but sent his invited paper, and H. Field, who attended the conference as a discussant and was invited by the editor to contribute his present paper. G. Jäger and H. Schwichtenberg were each joined by a co-author for their paper included here. A number of invited speakers delivered a talk, but were unable to contribute a paper to the volume because of other commitments; these are W. Buchholz, P. Hylton, H. Kamp, P. Martin-Löf, Y. Moschovakis, C. Parsons, G. Priest, and A. Urquhart.

The papers were all originally written for this volume, with two exceptions. The original place of publication of H. Field's paper "The Consistency of the Naive Theory of Properties" is *The Philosophical Quarterly*, volume 54, No. 214, January 2004. It is reprinted here by kind permission of Blackwell Publishing and the author. A condensed version of Hazen's paper appeared as A. P. Hazen and J. M. Davoren, Russell's 1925 logic, in: *Australasian Journal of Philosophy*, volume 78, No. 4, pp. 534–556, December 2000. Reprint of this material as part of the present paper by kind permission of Oxford University Press and the co-author.

The occasion to publish this volume would never have arisen without the essential financial support from a number of institutions that made the conference possible in the first place. It is therefore fitting to express my deep appreciation here to the *Deutsche Forschungsgemeinschaft* (*DFG*) for covering the main bulk of the conference costs through funds made available to the interdisciplinary DFG graduate program *Graduiertenkolleg "Sprache, Information, Logik"* (*SIL*) at Munich University, and special conference funds; furthermore to the following institutions and private sponsors for filling the remaining financial gaps: the *Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst*, the *Philosophy Department, University of Munich*, the *Gesellschaft von Freunden und Förderern der Universität München* (*Münchener Universitätsgesellschaft e.V.*), the *University of the German Federal Armed Forces Munich*, *Apple Computer Germany*, and *Alpina Burkard Bovensiepen GmbH & Co.*

Many people helped in various ways to organize the conference and to prepare this volume. I wish to thank all of them, in particular Solomon Feferman for his encouragement, and Andrew Irvine for generous advice, in the early stages of the conference project. Ulrich Albert carried the main organizational burden of the conference as "front liner" on the internet and at the conference site; Sebastian Paasch was responsible for the technical part of the preparation of the volume. My thanks to both of them for their efficiency and their commitment far beyond the call of duty. Ulrich and Sebastian were assisted at different stages by Marie-Nicole Ehlers, Martin Fischer, Roland Kastler, Uwe Lück, Michaela Paivarinta, Hannes Petermair, Marek Polanski, Christian Tapp, Mai-Lan Thai, Julia Zink, and Mauricio Zuluaga. Daniel Mook helped with checking the English of non-native speakers.

I owe a special debt to Karl-Georg Niebergall, who was always close at hand for any advise on matters of content pertaining to the contributions of the volume, and whose philosophical seriousness, enthusiasm and unfailing technical judgment I have had the privilege to enjoy in many years of joint seminars and private discussions.

Finally, I wish to extend my thanks to Dr. Manfred Karbe of Walter de Gruyter Publishers for supporting the decision to publish the volume, and for his sustained encouragement, advice, and patience during the phase of its preparation.

May 2004 G. L.

Table of Contents

Preface v

Godehard Link
Introduction. Bertrand Russell—The Invention of Mathematical Philosophy 1

W. Hugh Woodin
Set Theory after Russell: The Journey Back to Eden 29

Harvey M. Friedman
A Way Out 49

Sy D. Friedman
Completeness and Iteration in Modern Set Theory 85

Kai Hauser
Was sind und was sollen (neue) Axiome? 93

Gerhard Jäger and Dieter Probst
Iterating Σ Operations in Admissible Set Theory without Foundation:
A Further Aspect of Metapredicative Mahlo 119

Solomon Feferman
Typical Ambiguity: Trying to Have Your Cake and Eat It Too 135

Karl-Georg Niebergall
Is ZF Finitistically Reducible? 153

Tobias Hürter
Inconsistency in the Real World 181

Michael Rathjen
Predicativity, Circularity, and Anti-Foundation 191

John L. Bell
Russell's Paradox and Diagonalization in a Constructive Context 221

Peter Schuster and Helmut Schwichtenberg
Constructive Solutions of Continuous Equations 227

Kai F. Wehmeier
Russell's Paradox in Consistent Fragments of Frege's *Grundgesetze der Arithmetik* 247

Andrea Cantini
On a Russellian Paradox about Propositions and Truth 259

Hartry Field
The Consistency of the Naive Theory of Properties 285

Ulrich Blau
The Significance of the Largest and Smallest Numbers for the Oldest Paradoxes 311

Nicholas Griffin
The Prehistory of Russell's Paradox 349

Gregory Landini
Logicism's 'Insolubilia' and Their Solution by Russell's Substitutional Theory 373

Philippe de Rouilhan
Substitution and Types: Russell's Intermediate Theory 401

Francisco Rodríguez-Consuegra
Propositional Ontology and Logical Atomism 417

Bernard Linsky
Classes of Classes and Classes of Functions in *Principia Mathematica* 435

Allen P. Hazen
A "Constructive" Proper Extension of Ramified Type Theory (The Logic of *Principia Mathematica*, Second Edition, Appendix B) 449

Andrew D. Irvine
Russell on Method 481

Volker Peckhaus
Paradoxes in Göttingen 501

David Charles McCarty
David Hilbert and Paul du Bois-Reymond: Limits and Ideals 517

Jan Mycielski
Russell's Paradox and Hilbert's (much Forgotten) View of Set Theory 533

Shaughan Lavine
Objectivity: The Justification for Extrapolation 549

Geoffrey Hellman
Russell's Absolutism vs. (?) Structuralism 561

Robert S. D. Thomas
Mathematicians and Mathematical Objects 577

Holger Sturm
Russell's Paradox and Our Conception of Properties, or: Why Semantics Is no Proper Guide to the Nature of Properties 591

Vann McGee
The Many Lives of Ebenezer Wilkes Smith 611

Albert Visser
What Makes Expressions Meaningful? A Reflection on Contexts and Actions 625

List of Contributors 645

Name Index 649

Subject Index 653

Introduction. Bertrand Russell—The Invention of Mathematical Philosophy

Godehard Link

1. The Grand Conjunction

It is a theorem of elementary logic that, given an arbitrary two-place relation R, there is no object y such that for any x, x bears R to y just in case x does not bear R to itself. If the domain is the universe of sets, and R is the membership relation, then the theorem says there is no set consisting of those and only those sets that are not members of themselves. However, the 'naïve', unrestricted, comprehension scheme claims the contrary and thus produces Russell's paradox. By the same token, if R is the relation of predicability and the domain consists of predicates then there is no predicate that is predicable of just those predicates that are not predicable of themselves. A collection schema assuming the contrary yields the form of the paradox in which it was first formulated by Russell in May 1901. In the 1901 draft of Part I of his *Principles of Mathematics* he writes:

> The axiom that all referents with respect to a given relation form a class seems, however, to require some limitation, and that for the following reason. We saw that some predicates can be predicated of themselves. Consider now those … of which this is not the case. … [T]here is no predicate which attaches to all of them and to no other terms. For this predicate will either be predicable or not predicable of itself. If it is predicable of itself, it is one of those referents by relation to which it was defined, and therefore, in virtue of their definition, it is not predicable of itself. Conversely, if it is not predicable of itself, then again it is one of the said referents, of all of which (by hypotheses) it is predicable, and therefore again it is predicable of itself. This is a contradiction. ([70]: 195)

It is of some historical interest that Russell should first write down his paradox in the predicate version. According to his own account ([80]: 58) he had discovered it by analyzing Cantor's paradox of the greatest cardinal, where the context was set theory. Given that, it seems that Russell had already isolated the precarious scheme inviting paradox, viz., diagonalizing a relation and negating it, which is the common core of

the Liar, Cantor's Theorem, and Gödel's Incompleteness Theorem. As Gödel would put it years later:

> By analyzing the paradoxes to which Cantor's set theory had led, he freed them from all mathematical technicalities, thus bringing to light the amazing fact that our logical intuitions (i.e., intuitions concerning such notions as: truth, concept, being, class, etc.) are self-contradictory. ([28]: 131)

Of course, the historical record suggests that it took Russell some time to realize that he had discovered something that fundamental.[1] In particular, it was not before Frege's reaction of shock, a full year later, upon receiving Russell's famous letter describing the paradox—now also in its class form—that Russell came to appreciate its real impact and felt reassured enough to publish his findings. Thus the paradox first appeared in print with the publication of the *Principles* in May 1903.

According to existing reports the paradox was independently discovered by Ernst Zermelo. The first reference to this appears in a letter that Hilbert wrote to Frege in late 1903 thanking him for a copy of the second volume of the *Grundgesetze* which Frege had sent to him. "The example you give at the end of the book," says Hilbert, referring to the afterword in which Frege acknowledges Russell's paradox, "has been known to us here," adding in a footnote: "I think it was 3-4 years ago that Dr Zermelo found it after I had communicated my [own] examples to him" ([24]: 24). It is likely that by his "[own] examples" Hilbert meant what has become known as "Hilbert's Paradox",[2] to which I return presently.

Given the small step from the proof idea of Cantor's Theorem to the idea of the Russell class it should come as no surprise that an astute mind like Zermelo hit upon it by himself. What might be surprising is that Zermelo didn't publish it. Before arriving in Göttingen—and this is less known—he had already left a mark in the history of science in quite a different field. As an assistant of Max Planck in Berlin he had, drawing on a theorem of Poincaré's, devised what is called the *recurrence paradox*, which together with Loschmidt's reversibility argument put a serious threat to Boltzmann's program of reducing thermodynamics to statistical mechanics.[3] So why didn't he 'strike' again, this time threatening Cantor's newly erected edifice of set theory? The usual speculation about this is that at the time "the example" wasn't taken seriously by him, nor by Hilbert, for that matter.[4] Most likely, Zermelo, who had become a convinced Cantorian, viewed the problem as belonging to the same category as Cantor's paradox of the greatest cardinal and Burali-Forti's problem of the greatest ordinal. Presumably, these were all problems of a 'regional' kind, rather to

[1] See Griffin's contribution in the present volume.

[2] See [57], and also Peckhaus's paper in this volume.

[3] The father of quantum theory seemed to be impressed by it, which might have been part of the reason why he remained skeptical about Boltzmann's atomism for such a long time; see, e.g., [46]: 26f.

[4] For instance, assuming the paradox was discovered and communicated to him before the Paris meeting in August 1900, Hilbert could have mentioned it there in his famous lecture on mathematical problems.

be played down than exaggerated,[5] having to do with extrapolating arithmetic into the transfinite, and living within the bounds, or even on the fringes, of Cantor's theory; or so it seemed.

By the time he wrote his letter to Frege, however, Hilbert at least must have changed his attitude regarding these anomalies. For there he called his own examples "even more convincing contradictions" ([24]: 24), more convincing because he considered them to be of a "purely mathematical nature."[6] He continues with the explicit assessment that there is something seriously amiss in the very foundations of logic and mathematics. So the diagnosis was there now, and it was the same that the philosopher Russell had arrived at. However, the cure was not to come for years. As for Hilbert's part, the main focus of attention at the time lay elsewhere, on classical number theory, culminating in his proof of Waring's theorem in 1908, on integral equations and on mathematical physics, activities intensified by his close cooperation with H. Minkowski, up to the latter's premature death in 1909 (see, e.g., [68]). Thus in Göttingen, progress in the foundational work was basically left to Zermelo, who, however, had a different, much more mathematical, perspective, being preoccupied with the axiom of choice and its justification. While he did lay the foundations of modern set theory in the course of it, these circumstances nonetheless gave Russell a start of a number of years which he needed, after many abortive attempts full of blind alleys and frustrating toil, to find his own way out of the paradoxes. By bringing all the 'regional' versions of the paradoxes together, including the 'semantical' ones, he was able to broaden the scope of the investigation and to contribute to the foundations of logic proper.

Paradoxes like the Liar or those about the nature of infinity had been around as famous puzzles since antiquity, but the tools to tackle them had not been forthcoming in all that time. What counted as logical discourse, in Cambridge and elsewhere[7] was replete with ambiguity and obscurity. Precision and rigor were missing, and traditional logic had no way to come to serious terms with such genuinely logical problems as the paradoxes, old and new.

It is a curious fact about the history of knowledge that early on, at the time of the Greeks, logic should have dissociated itself from mathematics. Perhaps this is simply due to the contingent fact that the man who invented formal logic, Aristotle, was for all his greatness not known for excelling in the field of mathematics. At that time

[5] Those earlier problems didn't even seem paradoxical to Cantor and Burali-Forti, as G. Moore [54] argues.

[6] See ([57]: 168), for the phrase used by Hilbert. The authors explain that Hilbert's construction doesn't make use of the particular rules of ordinal and cardinal arithmetic set up by Cantor. Also, the editors of [24] add a footnote to the passage quoted in which they refer in turn to a remark by O. Blumenthal in his biographical essay of Hilbert [6]. There Blumenthal speaks of "the example, devised by Hilbert and nowhere leaving the domain of purely mathematical operations, of the contradictory set of all sets formed by union and self-mapping [Selbstbelegung]" ([6]: 422).

[7] The work of the neo-Hegelian philosopher F. Bradley, for instance, marks a point of reference for the appreciation of the tremendous change that the subject of logic was to undergo in the course of Russell's contribution.

began what I like to call the Babylonian captivity of logic in the realm of metaphysics, lasting for over 2000 years. When in the age of the scientific revolution Aristotelian metaphysics became the target of modernizers like Galileo, logic was considered part and parcel of metaphysics and was dismissed together with the philosophy Galileo fought against. For him, formalization of logic was obsolete; what was needed from logic he considered as natural and no real subject of study, certainly no precondition for founding the new science of physics. In his polemic *The Assayer* [25] Galileo said famously:

> [Natural] philosophy is written in this grand book, the universe, which stands continuously open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it. ([25]: 237f.)

Thus, if even syntax was in the domain of geometry, little room was left for logic. In fact, the subject of logic was not only ignored, it acquired an outright bad reputation with scientists and mathematicians over the centuries, still to be felt in Poincaré's polemics. Even Kant, while certainly not prejudiced against logic, said nonetheless that logic hadn't made any progress since Aristotle. He saw this subject as a beautiful finished edifice which could only suffer from further additions to it:

> If some of the moderns have thought to enlarge it [logic] by introducing *psychological* chapters on the different faculties of knowledge (imagination, wit, etc.), *metaphysical* chapters on the origin of knowledge or on the different kind of certainty … or *anthropological* chapters on prejudices, their causes and remedies, this could only arise from their ignorance of the peculiar nature of logical science. ([44]: Preface, B VIII)

Here Kant should not be taken to mean that logic cannot be applied to other fields of knowledge, but rather that the subject matter of logic proper is neither psychology nor metaphysics. But what Kant complained about had happened all the time and was soon to culminate prominently again in Hegel's work and that of his followers. This was also the situation in England when Russell arrived at the scene.

So what was Russell's point of departure? In hindsight, it can be said that Russell didn't discover his paradox quite by accident. The late 1890s saw him deeply involved in studying what was known among idealist philosophers as the 'contradictions' of the infinitely large and the infinitely small. At the time, philosophy in Cambridge was dominated by neo-Hegelians, and Russell adopted their point of view.[8] Thus, in spite of his mathematical training, he took a philosophical interest in mathematics and its foundations. Even after he had given up his idealist convictions he kept his thoroughly

[8] See, for instance, [41], [37], [54].

philosophical outlook. His ontology had become pluralistic, but the 'theoretical terms' of his logic consisted of the same kind of philosophical entities that traditional logicians like Bradley would use, viz., propositions, concepts, and, with a novel emphasis on their importance, relations. In contrast, for instance, reflection on language as a prerequisite to modern logic played no particular role yet in Russell's work. The *Principles*, even in those parts where he introduced the innovation of the denoting relation for the first time, were written in a rather traditional philosophical style, certainly not apt to draw a large readership among mathematicians.[9] The situation was just like Frege described it in the Preface of his *Grundgesetze* [23], where he said that he had to give up on the mathematicians who would call his work "metaphysics" because of expressions like 'concept', 'relation', 'judgement' occurring in it. There were those two separate communities, and it seemed hard to imagine how they should take interest into one another. Mainstream mathematicians did not only care little about philosophy but considered even Cantor, who was really one of them, as too philosophical to be taken seriously. It is a telling historical detail that Felix Klein's widely read lectures on the development of mathematics in the nineteenth century [45] mentions Cantor only in passing. Russell clearly expressed this state of affairs in *My Philosophical Development* [80], where he says:

> The division of universities into faculties is, I suppose, necessary, but it has had some very unfortunate consequences. Logic, being considered to be a branch of philosophy and having been treated by Aristotle, has been considered to be a subject only to be treated by those who are proficient in Greek. Mathematics, as a consequence, has only been treated by those who knew no logic. From the time of Aristotle and Euclid to the present century, this divorce has been disastrous. ([80]: 51)

Here is now what I consider Russell's historic achievement. Open-minded, quick, and polyglot, he was able to engage philosophers and mathematicians alike in an excitingly new foundational debate. And crucially, this was not just another epistemological enterprise, important as meta-reflections are for clarifying the conceptual and methodological premises of a given discipline. It was all that, but it was more: it opened up new horizons for technical research of an unprecedented kind, deriving from a few basic logical premises. Russell, who was surely not equipped with the most advanced mathematical expertise of his time,[10] nevertheless almost single-handedly produced viable solutions to the new problems that were rigorous enough to spark interest in the mathematicians' camp, in particular, with Poincaré and the Hilbert school. Poincaré, highly skeptical about symbolic logic, still had a hand in uncovering the presuppositions of quantification, while Hilbert was of course the one influential mathematician who recognized early on that the paradoxes did pose a problem for

[9] For instance, Peckhaus [this volume] cites G. Hessenberg in Göttingen for "scathing" comments on Russell's book.

[10] Mathematics at Cambridge was not the best to be had in Europe at the time, as Russell admits ([80]: 29).

the foundations of mathematics, and that something had to be done about it.[11] Some of the finest younger mathematicians who came to specialize in the new discipline or at least substantially contributed to it at some point in time (Weyl, Bernays, von Neumann, Gödel) took their starting point from the framework of *Principia* in one way or the other. On the philosophers' side, Wittgenstein, Ramsey, Carnap, and Quine stand out; each of them adopted the Russellian outlook as well, although Wittgenstein, of course, was soon to strike out in quite a different direction.

It is this combination of transdisciplinary philosophical query, mathematical training and logical expertise that characterize Russell's paradigm of mathematical philosophy. Russell therefore occupied a unique role in bringing about what I like to call, in an astronomical metaphor, the *Grand Conjunction* of the old and venerable fields of philosophy, logic, and mathematics.

Let us remind ourselves of some important features of Russell's work that are representative for this new enterprise.

1.1. Some Russellian Themes

1.1.1. Symbolic Explicitness

Today it is hardly necessary to stress the benefits of symbolism, and in particular, the salutary role of a good symbolism. The Peano–Russell notation of *Principia* was indeed a good one, serving a generation of logicians well (Frege, for instance, was less lucky in convincing others to use his symbolism, in spite of his clarity of insight and meticulous exposition). At first, when Russell seized upon Peano's notation, it was perhaps not more than an expedient to express ideas succinctly; historically, however, it is more than that. It was a unique way, in philosophy and even in mathematics, of 'making it explicit,' to use a catch phrase of today. Distinctions could be drawn that wouldn't be noticed otherwise, thus preparing the ground for new questions and lines of research. The organizing power of the symbolism made Russell become definite about his primitives, aware of what could be defined in terms of what, and what was needed to prove an assertion. In his case in particular, it also led to productive strictures with the conceptual apparatus he started out with. This is most clearly seen in his gradual retreat from the received metaphysical approach towards his fundamental innovations, as documented in his writings from the *Principles* leading all the way up to *Principia*. In the latter work, he finally was in the position to give proofs and derivations from first principles, *more geometrico*. It strikes the eye that he had thereby left behind everything that was so far considered 'logic' in the philosophers' sense.

Even so, we should not pass over the weak points that could not be covered up by the extensive formalism; as Gödel said: "It is to be regretted that this first comprehensive and thorough going presentation of a mathematical logic and the derivation

[11] See again Peckhaus's paper in this volume with references therein, and also [50].

of Mathematics from it is so greatly lacking in formal precision in the foundations (contained in *1–*21 of *Principia*), that it presents in this respect a considerable step backwards as compared with Frege" ([28]: 126). Also, what we today take as the hallmark of modern mathematical logic, a notion of *metatheory*, was barely to be found (but see below). Related to this failure is the notorious laxity in matters of use and mention, which Quine famously harped on and induced him to basically rewrite *Principia* from scratch. Thus I think it is fair to say that Russell's logic was still in a pre-paradigmatic stage of development. The real 'founding document' of the new discipline was Gödel's paper on formally undecidable propositions [27]; but as its complete title shows, the point of reference were the *Principia Mathematica* of Whitehead and Russell.

1.1.2. On the Road from Metaphysics to Logical Syntax

In his unique laboratory of ideas reflected in the *Principles*,[12] Russell worked with traditional philosophical entities like propositions and concepts. Propositions were certain complexes of 'terms', that is, objects,[13] which—though clearly patterned after their syntactic form—were no syntactic entities. Rather, they were conceived in analogy with regular physical objects consisting of various extra-linguistic parts. The notion of 'constituent' used in this connection was ambiguous, meaning both a part of an expression and a part of complex entity. Real syntactic leverage is introduced, however, when Russell begins to inquire into the notion of *substitution*, with which he experimented a lot over the years. This investigation had the natural tendency to acquire a syntactic character almost by itself. But for Russell the transition to syntax was by no means automatic. Quite to the contrary, he struggled with the formalism and its interpretation for a long time, trying to do equal justice to the demands of metaphysics and the emerging discipline of symbolic logic. This is the kind of tension mentioned above that led to all the endemic ambiguities that modern readers of Russell justly complain about,[14] wondering time and again whether Russell is talking about entities or expressions. I think that the failure to distinguish clearly between the expression and the object it signifies stems from a *neglect of language* in Russell's early work where it is not expressions but concepts that denote ([71]: chap. V). Philosophically, his own 'linguistic turn' was still to come, in spite of all the symbolic machinery he had already adopted.[15] Even in his seminal essay *On Denoting*

[12] As Quine says of this book: "It is the more remarkable, then, that this prelogical logic of Russell's, this morass of half-formulated problems, should already contain the embryo of twentieth-century philosophy" ([62]: 5).

[13] Recall that the word 'term' in the *Principles* means an entity and not an expression.

[14] For instance, even in *Principia* he could still write: "A function, in fact, is not a definite object, … it is a mere ambiguity awaiting determination" ([85]: 48).

[15] This judgement seems to contradict what Russell himself says about that time: "I was very much occupied, in the early days of developing the new philosophy, by questions which were largely linguistic. I was concerned with what makes the unity of a complex, and, more especially, the unity of a sentence" ([80]: 49). However, the very example he gives—involving the philosophically charged concept of part-whole—is an indication of the fact that it is still mainly metaphysics that he worries about.

[72] we find him wavering between the newly introduced 'denoting phrases' and the old 'denoting complexes'. The infamously obscure *Gray's Elegy* argument occurring there bears ample witness of the fact that Russell had not quite worked out the basic semiotic relations.[16]

The other principal case is point is the ambiguous status of the propositional functions, which, after Russell had abandoned classes as regular objects, had to carry the main ontological burden of *Principia*'s philosophy. The exact criterion of identity of propositional functions—or of some equivalent intensional notion of property—has remained a matter of debate up to the present day, that is, whether they should be conceived as equipped with outright syntactic granularity or something more coarse-grained yet short of extensionality.

When I speak of Russell's road to syntax I don't mean to say that he left metaphysics behind completely; not so, as his theory of logical atomism shows. But in his hands metaphysics or ontology became more amenable to a logical treatment in the modern sense. The development also showed, however, how problematic the old metaphysical conception of ontology became which still constituted the philosophical underpinning of *Principia*. Russell's intellectual crisis in 1913, produced by the events surrounding the unfinished *Theory of Knowledge* manuscript [75], testifies to this; however great the role of Wittgenstein's criticisms was in "paralyzing" Russell, the internal problems of traditional metaphysics still occupying his mind were such that they led the whole program of logical atomism into an impasse.[17]

1.1.3. The Re-Invention of Ockham's Razor As a Logical Tool

This is, of course, Russell's famous method of eliminating dubious entities by the technique of incomplete symbols in his theory of description, first laid down in *On Denoting* and elaborated in *14 of *Principia*. It not only swept away the notorious round square and all its lookalikes, but also expelled classes as first-class citizens from his ontology. Now it is one thing to proclaim that entities should not be multiplied beyond necessity, and quite another to devise an impeccable logical method of filling this slogan with meaningful content. That this was not an altogether trivial matter is shown by the complication arising from the presence of, first, logical operators, which are sensitive to scope, and secondly, more than one descriptive phrase, where the order of elimination becomes an issue. Russell provided viable solutions to these problems, albeit refined by others later on in various directions. Thus, his theory of descriptions is rightly hailed as a landmark of genuine logical philosophy.

The advent of this method marked a decisive point on Russell's road to logical syntax. The denoting concepts of the *Principles* evolved into denoting phrases and dissolved altogether as an ontological substance. Incomplete symbols are plainly

[16] However, Landini's paper in this volume contains an intriguing case for the claim that in spite of all this, Russell's "substitutional theory" can be reconstructed rather faithfully in a way as to contain a viable solution to the paradoxes and a surprising justification of logicism *prior* to *Principia*.

[17] On this, see also Rodríguez-Consuegra's paper in this volume.

expressions, there are no 'incomplete entities' to be had. Incompleteness in this sense derives from the semiotic function attached to linguistic expressions. Not every part of a meaningful expression needs to be meaningful, some parts may require a *context* to become so. This notion of context doesn't make sense with extra-linguistic objects, which (usually) lack a semiotic role. In the special case of mathematical objects, it is true that Frege called functions "incomplete" or "unsaturated"; but we mustn't take that literally here: in fact, Frege first models logic after mathematics by assimilating concepts to mathematical functions, and then uses the syntactic (indeed, chemical) metaphor of saturation to characterize functions again.

In the foundations of mathematics, the main benefit Russell derived from the technique of incomplete symbols was, of course, the prospect of doing mathematics without a commitment to classes.

> The following theory of classes, although it provides a notation to represent them, avoids the assumption that there are such things as classes. This it does by merely defining propositions in whose expression the symbols representing classes occur, just as, in *14, we defined propositions containing descriptions. … The incomplete symbols which take the place of classes serve the purpose of technically providing something identical in the case of two functions having the same extension. ([85]: 187)

In view of current discussions about anti-Platonist approaches to an understanding of mathematical practice[18] we have to remember that Russell's *no-classes theory* didn't spring from any nominalist convictions—quite the opposite, he was a realist regarding universals, in particular, relations. The Quinean extensionalist reform, which basically equated all *universalia* with classes, was still to come, so ousting classes was not a nominalist project for Russell. Rather, it seems that the effects the paradoxes had on him were such that he never trusted classes again, which he took responsible for engendering the antinomies. By reducing class talk to propositional function talk Russell paved the way for a *conceptualist approach* to set theory.

1.1.4. The Nature of Quantification

As far as quantification is concerned Russell had to come a particularly long way from where he started out. For instance, in his *Principles* he cannot make up his mind about the semantic mechanism involved in the assertion of a noun phrase like *some man*. He wonders "whether an ambiguous object is unambiguously denoted, or a definite object ambiguously denoted" ([71]: 62). To save his newly discovered denoting relation from this ambiguity he shifts it to the object denoted, with obviously little convincing results. Metaphysical constraints about what can or cannot not be part of the proposition expressed by an assertion involving phrases like *some man* or *all men* make him come up with an account that is squarely at variance with the Frege–Peano style of quantification which Russell uses elsewhere in the same book.

[18] See, e.g., [21], [40].

But there is a piece of *linguistic philosophy* emerging, which Russell puts to logical use, when he draws a distinction between the genuinely universal quantifier *all* and the logical meaning of free-choice *any*. This distinction, which appears in [74] and made it into the first edition of *Principia*, can be viewed as an instance of Russell stepping out of his single, all-encompassing formal system and moving towards some sort of metalanguage.[19] In [74], after reviewing the paradoxes, he concludes that they all have in common "the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself. This leads us to the rule: 'Whatever involves *all* of a collection must not be one of the collection' " ([74]: 63). This is one form of the Vicious Circle Principle, which however, as Russell goes on to say, makes "fundamental principles of logic" meaningless, such as 'All propositions are either true or false'. Here he finds a role to play for *any*: "Hence the fundamental laws of logic can be stated concerning *any* proposition, though we cannot significantly say that they hold of *all* propositions" ([74]: 68). This is because in Russell's type theory, the logical tool of overt quantification is, by necessity, *parametric*, i.e., relative to a given type; truly universal claims can only be made by a quantifier-free, *schematic assertion*. That might be interpreted as foreshadowing the finitary assertions of Hilbert's metamathematics. However, Russell never made the step to distinguish between an object language and a metalanguage, and so he didn't really know what to do with this observation in technical terms. Consequently, the separate role of *any* was given up in the second edition of *Principia*. It seems to me, though, that what Russell hinted at comes closer to what has been reconstructed in recent years as the concept of *arbitrary object*,[20] which is a philosophical idea.

Suffice it to say, then, that Russell did arrive at the modern notion of a quantified sentence as a *separate form*, equipped with a specific range for the bound ("apparent") variable. But, as Goldfarb [35] rightly points out, since there was no notion of *model*, in which the quantifiers and some non-logical vocabulary could be variously interpreted, neither he nor Frege, for that matter, can be said to have developed the full-fletched concept and machinery of current quantification.

1.1.5. The Hierarchy of Types

It was suggested above that it took Russell's original logico-philosophical frame of mind not only to recognize the serious nature of the problem of the paradoxes but to sit down and work out a 'non-regional', comprehensive solution as well. Putting the paradoxes in a proper philosophical perspective Russell also tried to include the so-called semantical paradoxes into his solution. Having absorbed Cantor's work, and with Peano's symbolic logic at his disposal, he now felt that he finally had in his hands the technical means to deal with the whole array of paradoxical phenomena besetting the foundations of logic and mathematics. The final result he came up with was the *theory of ramified types*, RTT for short ([74], [85]). RTT consists of two main

[19] This was noticed by Hazen [38].

[20] See Kit Fine's [22].

components, a simple theory of finite types, STT, and the device of ramification. STT is designed to take care of the set-theoretic paradoxes, while ramification is a way to resolve the semantical paradoxes.

The basic idea of STT is simple enough. It derives from a platonist ontology of individuals and properties (concepts, propositional functions), regimented into levels, in which not only individuals fall under properties but properties in turn under other properties, provided the subsuming property is always one level higher up. Assuming individuals to have type 0, we get properties of individuals as type 1 properties, properties of type 1 properties as type 2 properties, etc.; thus, the above concept of self-predicability cannot be expressed anymore. There is a *syntactic ban* on reflexive predication, negated or not. The same is true, derivatively, for classes, blocking Russell's paradox.

Virtually all of the usual mathematics (except set theory proper, of course) can be carried out in the framework of simple type theory; indeed, only a few types will be actually used to ascend from the natural numbers, which can be either given as individuals or constructed as second-order properties in the Frege-Russell style,[21] to the higher systems of the rational and real numbers, and on to real-valued functions, function spaces, etc.

However, not STT, but Zermelo's system Z of set theory (plus the axiom of choice, and strengthened by Fraenkel's replacement axiom) carried the day as the standard framework for mainstream mathematics. This was inevitable; mathematics had been moving towards a unification of its disciplines, but Bourbaki style, not for the price of being put in a syntactic straitjacket. In standard set theory, the types can easily be retrieved from the notion of rank, and the axiom of foundation blocks $\in$-cycles. Moreover, while Russell's logicist conception assimilates membership to predication, with its characteristic trait of non-transitivity, membership in set theory is relational from the outset, allowing for, and indeed making essential use of, transitivity in modelling, for instance, the well-ordering on the ordinals by the $\in$-relation. The predicational view precluded that, and with it the related idea of cumulativity, which was to become the central idea of the Zermelo–Gödel hierarchy of sets, typically pictured as the 'funnel of sets'. Iteration along transfinite types was essential for this hierarchy, whereas in the predicational picture, Russell could see no use for transfinite types. These were the main limitations of the non-ramified part of Russell's theory of types.

While type theory did not become the general framework for everyday mathematics it did survive in important quarters of specialization, in particular, foundational studies, recursion theory, and an array of applications, from computer science (see,

[21] Of course, the infinitely many objects the logicist needs to construct in order to be able to interpret arithmetic don't come for free. Russell saw no way of avoiding an axiom of infinity (but see Landini's paper in this volume for an argument to the contrary, at least as far as Russell's substitutional theory is concerned). In contrast, the (consistent) system of second-order logic with 'Hume's Principle' added, which has been called 'Frege Arithmetic' [5], does produce the infinite number series; isn't the neo-logicist thereby making good on his claim after all that arithmetic is logic? I don't think so. Hume's Principle adds 'ideology' (in Quine's sense) to second-order logic, by introducing the operator 'the number of . . .'. Even Frege himself didn't seem to regard Hume's Principle as a primitive truth of logic; see ([39]: 286).

e.g., [10], [3]) and artificial intelligence to categorial grammar theory ([2], [55]) to higher-order intensional logic ([53], [26]). The foundational work was mainly concerned with justification; in the course of these developments, the original setup was either simply used, as, e.g., in Robinson's non-standard analysis [69], or modified, either by the introduction of functional types[22] or more recently through amendments towards greater flexibility.[23] There is also the influential Martin-Löf style intuitionistic type theory [51], important both for its underlying philosophy and its range of applications.

1.1.6. Predicativism

This issue is a paradigm case of the kind of intimate connection between logic and philosophy described above. Brought into the discussion by Poincaré, the idea of a predicative definition was taken up by Russell and turned into a technical tool for dealing with the semantical side of the paradoxes, that is, the ramification of the type hierarchy.[24] It was only natural for Russell, who had come to deny the existence of classes beyond a convenient notation, to adopt a constructivist attitude towards propositional functions, which served as non-extensional proxies for classes. But then, classes depended on the definitional history of the propositional functions which gave rise to them. In particular, no propositional function could be introduced by referring to a totality to which it already, in an intuitive sense, belonged. This is the Vicious Circle Principle; Russell's formulation of it in [74], quoted above, is carried over to [85], where it is paraphrased as "If, provided a certain collection had a total, if would have members only definable in terms of that total, then the said collection has no total" ([85]: 37). Now Gödel ([28]: 135) pointed out that these two versions are not really synonymous, and that not 'involvement', but rather 'definability only in terms of' is the critical notion. It is the Vicious Circle Principle in this latter sense that found its technical expression in the predicative comprehension principles of *Principia* and, later on, of the standard systems like NBG set theory or the fragment ACA_0 of second-order arithmetic. According to Gödel, the principle applies

> … only if the entities involved are constructed by ourselves. … If, however, it is a question of objects that exist independently of our constructions, there is nothing in the least absurd in the existence of totalities containing members which can be described (i.e., uniquely characterized) only by reference to this totality.
> ([28]: 136; two footnotes omitted)

This is Gödel's well-known pronouncement of his platonist convictions. Thus the issue of predicativity brought into focus two major and opposing philosophies regarding mathematical objects: platonic realism and constructivism. They have been

[22] Beginning with Hilbert's and Ackermann's effort in the twenties to Church's lambda calculus to Gödel's famous *Dialectica* interpretation.

[23] See especially Feferman's [12], and up to the system W of [14].

[24] For an comprehensive overview of predicative logics and their philosophy, see [38].

described at length in the literature, and need not be rehearsed in detail here. But the interesting thing is that there are two historical developments deriving from one and the same source, while conflicting in their philosophies: On the one hand, there is the predicativist tradition starting with Hermann Weyl and leading up to the modern field of predicative proof theory initiated by Feferman and Schütte; but on the other hand there is Gödel who transformed the idea of ramification in such a way as to produce the first model of Zermelo–Fraenkel set theory (given ZF is consistent), the constructible universe. He achieved this by adopting an explicit non-constructivist attitude towards the objects of set theory (or towards the ordinals anyway, working constructively from there). Another often quoted remark of Gödel's, in a letter to Hao Wang, makes this clear: "However, as far as, in particular, the continuum hypothesis is concerned, there was a special obstacle which *really* made it *practically impossible* for constructivists to discover my consistency proof. It is the fact that the ramified hierarchy, which had been invented *expressly for constructivistic purposes,* has to be used in an *entirely non-constructivistic way*" ([34]: 404).

I'd like to make two points here. The first concerns the relationship, in the history of the exact sciences, between the genesis of new results and the underlying philosophical outlook (if any) that influences them. When new ideas possess at least some degree of technical explicitness they can be exploited and recombined in quite productive and unforeseen directions regardless of their original purpose or philosophy. In the case of Russell's ramified type theory this is one of the benefits flowing from the effort made in *Principia*, quite independently of its subsequent fate as foundational system. Secondly, the case shows the futility of constructing continuous genealogies in the history of ideas. In my view, there is little substance in claims of the form, "It was all in X", where 'X' stands for 'Russell', 'Frege', or whoever the favorite hero of an historiographer happens to be. For instance, in spite of the credit that Gödel himself gave Russell regarding the sources of the constructible universe, we would rather agree with R. Solovay, the commentator of pertinent writings of Gödel in the *Collected Works*, who says: "There seems to me to be a vital distinction between the precise notion of Gödel and the somewhat vaguer discussions of the ramified hierarchy found in Russell's writings. Thus Gödel's well-known comments … to the effect that his notion of constructibility may be regarded as a natural extension of Russell's ramified hierarchy into the transfinite now strikes this writer as much too generous" ([83]: 120). There was indeed both a decisive conceptual break and a new level of technical sophistication that distinguishes Russell's notion of ramification from what it became in Gödel's hands. Even so, the historical importance of Russell's efforts is not diminished by the fact that they were superseded by later developments; rather it should be measured by its potential for generating fruitful lines of research.

Let me shortly mention two more such lines of research that can be traced back to Russell's predicativism. One is the theory of truth; comparing Russell's approach to the semantical paradoxes with that of Tarski, A. Church concludes: "Russell's resolution of the semantical antinomies is not a different one than Tarski's but is a special case of it" ([9]: 301). Thus Tarski can be said to have extended the ramified hierarchy

in his theory of truth, albeit introducing essential innovations in the course of it. The other development, already mentioned above, concerns the modern research program of predicative mathematics. It finds its crystallization in a recent work of S. Feferman's [14] which takes Weyl's predicative program in *Das Kontinuum* and moulds it into a rigorous theory, called W, of 'flexible' finite types. There is a proof, jointly due to Feferman and Jäger ([17], [18]), that W is both a conservative extension of, and proof-theoretically reducible to, Peano Arithmetic. Thus W is drastically weaker than any theory of the order of ZFC. Yet Feferman claims that basically all scientifically applicable mathematics could be formalized in W.[25] This is a much more specific argument than those usually given in the debate on the indispensability of mathematics for the sciences initiated by Quine and Putnam. In fact, Quine acknowledges Feferman's program, calling it "a momentous result", if realizable: "It would make a clean sweep of the indenumerable infinites and unspecifiable sets" ([64]: 230). For further discussion of this, see [15].

1.1.7. Reductionism

We have already touched upon major reductionist issues in Russell's work which met with some success, like the elimination of definite descriptions, the reduction of classes to propositional functions, and the predicative reform of logic. However, the main logicist research program Russell embarked on, the reduction of mathematics to logic, turned out to be a failure. The principal reason was, of course, that substantial existence assumptions needed in mathematics (i.e., axioms of infinity) are alien to the province of pure logic.[26] But the program also foundered on technical problems, like those related to the infamous *axiom of reducibility*.[27] Also, the extension to epistemology of the reductionist method, hailed by Russell as the "supreme maxim in scientific philosophising" [76], and followed by Carnap in his *Logischer Aufbau der Welt*, did not succeed; for many respectable reasons it was abandoned in the course of the last century.[28]

[25] See also the postscript to [14] in [16], 281–83.

[26] It is a curious fact that Russell never seemed to have acknowledged this failure. As his discussion in [78], for instance, clearly shows, he is fully aware of the fact that the Peano axioms have no finite models, and that the injectivity of the successor function cannot be proved from pure logic. In fact, he calls the axiom of infinity a "hypothesis", like the multiplicative axiom (choice) and the axiom of reducibility, which he doesn't consider logically necessary. Yet in the same book he squarely equates mathematics with logic: "Pure logic, and pure mathematics (which is the same thing), aims at being true, in Leibnizian phraseology, in all possible worlds, not only in this higgledy-piggledy job-lot of a world in which chance has imprisoned us" ([78]: 192).

[27] See, e.g., [38] for the relevant background, but also [9] for the claim that, contrary to what has been common coin about this ever since Ramsey's criticism, repeated by Quine, the axiom by no means defeats Russell's purpose of ramification if only the essentially intensional character of *Principia*'s logic is taken into account. As far as the last point is concerned, Linsky's paper in this volume is also relevant here.

[28] Basically, a theory of the physical world just cannot be a definitional extension of a phenomenalist theory of sense-data. For a recent discussion of the philosophical flaws of the phenomenalist viewpoint see, e.g., [60].

Yet in the domain of logic and the philosophy of mathematics, the reductionist spirit has survived and even grown stronger as more sophisticated logical tools became available, like the concepts of relative interpretation between theories, conservativity of one theory over another, or proof-theoretic reduction. This line of research flowed from two main sources, Hilbert's metamathematics and the nominalist tradition in modern logical philosophy.

As for the former, it is true that Hilbert's original program of finitary consistency proofs was given up in the aftermath of Gödel's incompleteness theorems; but modern proof theory kept the basic methodology while relaxing finitist requirements [30], calibrating theories according to their relative consistency strength, investigating reductive relations between them [13], and quite generally exploring the wealth of subsystems of analysis, i.e., of second order number theory, for the development of most of everyday mathematics. In addition to the predicativist research program mentioned above there is the important program of "Reverse Mathematics" [82] singling out various set existence principles which are not only sufficient but also necessary for specific portions of positive mathematics, thereby uncovering the precise resources used.

The other major source of reductionist philosophy of mathematics is the nominalist program initiated by Goodman and Quine [36]. While Quine in his famous turn [61] embraced classes as indispensible for scientific practice, H. Field [21] tried to justify this practice by showing its conservativity over a nominalist ground theory. Yet another approach is S. Lavine's intriguing "theory of zillions" [48], which is committed to the justification of unrestrained impredicative set theory on the basis of the resources of "finite mathematics" present in J. Mycielski's concept of locally finite theories [56].

The leading rationale common to all these approaches is a kind of reasoning that is "conscious of its resources" [49]. The typical question would not only be, "Can we prove it?", but more specifically, "In which theory can we prove it?", or "What is the weakest theory for establishing the result or, more generally, for justifying usual mathematical practice?" Note that from a foundational point of view the question of accepting a theory is thereby 'factorized' into two component steps: (i) the technical question as to what the necessary principles are which the theory rests on, and (ii) the (mainly) philosophy-driven decision to embrace or repudiate the principles and thereby the theory.

According to the received view, first principles ("axioms") have to be not only true but self-evident. But as A. Irvine's paper in the present volume reminds us, there is an element in Russell's methodology, called the "regressive method" [73], which calls for admitting principles that might suggest themselves as axioms only after a considerable amount of technical work. Thus, the above decision (ii) may after all not be grounded—or at least not exclusively grounded—in some philosophical predilection or other, a situation definitely to be welcomed. It is interesting to note that Gödel was aware of this feature in Russell's work and supported it.[29]

[29] Remarks that make direct reference to Russell can be found in ([28]: 127f.); see also his well-known statement on the question of new axioms in set theory, in ([29]: 521).

Now there is a striking analogy here to a methodological maxim in the otherwise uncompromisingly mathematical work of H. Woodin, expressed for instance in the popular [86] and also in his contribution to this volume. Woodin's paradigm case for mathematical truths that might be discovered only *a posteriori* is descriptive set theory. This field of research, which amounts to the definability theory of the continuum, saw a period of rapid progress in Russell's time through the work of the French school of analysts led by E. Borel, R. Baire, and H. Lebesgue, and the Russian group round N. Luzin, but then reached a stalemate when it met with independence phenomena, unrecognized as such at the time.[30] There is a hierarchy of complexity, very similar to the analytic hierarchy in classical recursion theory, in the subsets of the real numbers called the hierarchy of the projective sets. Central regularity properties for these sets, like the measure-theoretic question of Lebesgue measurability, turned out to be undecidable on the basis of the usual axioms of set theory already after a few stages up the hierarchy. In recent decades a group of so-called Determinacy Axioms were isolated emerging from the investigation of infinite games. It could be shown that the axiom geared to the projective sets, *projective determinacy*, settles their regularity properties in the context of ZFC, thereby providing a remarkably closed body of mathematical knowledge. Woodin says about this axiom:

> It is interesting to note that the understanding of this axiom took quite a number of years of research and that the credible claim for the truth of the axiom was only possible after this research was accomplished—there is no known elementary argument for the truth of the axiom. This fact creates something of a challenge for the skeptic to explain. ([88]: 31)

Woodin also gives us his opinion on the 'degree of truth' of the axiom *vis-à-vis* other more familiar axioms: "I believe the axiom of *Projective Determinacy* is as true as the axioms of Number Theory." (*ibid.*) This statement is of course patently at variance with the persuasions of many, including some authors of the present volume. It urges nothing else than that 'Cantor's paradise' of the actual infinite cannot easily be divided into an 'indispensible' and a 'recreational' (Quine [63]: 400) province.

1.2. Conclusion

What is the status of Russell's work and his project of mathematical philosophy a hundred years later? Russell's role in logic seems to me similar in many respects to Galileo's role in the field of physics. Today Galileo's law of falling bodies is derived in the first pages of any standard textbook of physics, just like Russell's paradox occurs on page 2 or so of treatises on set theory. Galileo was quite wrong in many important scientific issues of his time. For instance, rather than joining Kepler not only in his general Copernican outlook, but also in his bold hypothesis of the elliptical movement

[30] See [43] for a historical sketch of the problems involved.

of the planets, Galileo somewhat stubbornly stuck to the principle of circular movement of the heavenly bodies, blaming measurement errors for the observed deviations. A quarter of a century after his discovery of the *Law of Free Fall* he claimed that a stone dropped from a tower would describe a uniform linear movement superimposed on a likewise uniform circular movement.[31] Again, eager to ban reference to "obscure qualities" he came up with a faulty analysis of the tides, denying any influence of the moon. But Galileo is not only—and not even mainly—remembered for his simple law; rather, his deep structural insights into the nature of motion (relativity of motion, Galilei invariance), and the mathematization of natural philosophy in general, marked the beginning of the new field of physical science.

Russell's work can be viewed in quite a similar light. There are many points of technical detail, and of general methodology as well, which did not stand up to later developments. Yet his work changed the character of the traditional field of logic in a way that it was never the same again. The same holds true for philosophy. Russell had a lasting impact on the way to do philosophy, in spite of the fact that almost all of his positions were severely criticized later on, not only in non-analytic quarters but also under the influence of Wittgenstein, Russell's nemesis.

Mathematical philosophy in the Russellian style aspires to be informed by, and to reflect on, the latest and most advanced results in the fields of logic and mathematics,[32] and to inspire technical foundational work in turn. The centenary of Russell's discovery seemed a fitting occasion to document current activities in this field of interdisciplinary research.

2. The Contributions

The authors of this volume are mathematicians, logicians or philosophers and historians of logic and mathematics, all of them with a characteristic multiple expertise in more than one of these fields. Their contributions address different foundational issues from varying viewpoints, and in a different style. Some papers are quite technical, but even where the style is informal, the authors draw on solid technical knowledge. Thus the collection of papers constitutes a body of mathematical philosophy in the sense explained above. It also provides useful pointers to current developments in the foundations of mathematics, in many cases shaped to a large extent by authors of the present volume.

The contributions can roughly be divided into the following groups: (i) papers on current, post-Cohen era, set theory; (ii) proof-theoretic reflections on various non-standard approaches to set theory, and indeed on infinity itself; (iii) constructive mathematics; (iv) some modern technical answers to the paradoxes; (v) papers on Russell's

[31] This incoherent view was stressed by Feyerabend in his classic [20], where it is attributed to Galileo's 'tactics' of persuasion: the important point to get across was the earth's rotation, 'never mind the details'.

[32] And in the sciences in general, although this is not at issue in the present volume.

logical work; (vi) papers providing some additional historical context related to the foundational crisis; (vii) philosophical papers reflecting on issues raised by Russell's philosophical logic. Although there is always some arbitrariness in a linear order the papers are arranged following this subdivision, and within each group according to a thematic or historical vicinity.

The work of *W. H. Woodin* (see [43], [87]) has become known outside expert circles for its staggering attack on anti-Platonist agnosticism regarding the meaning of set theory, in particular for the claim, based on an impressive body of deep original mathematics, that Cantor's continuum problem is a meaningful question after all, and that it should be settled in the negative. His paper *Set Theory After Russell. The Journey Back to Eden* contains a succinct recapitulation of the main line of argument, together with some new results mentioned and an exposition of the central notion of Ω-logic which differs from earlier accounts. — One of the major themes in *H. Friedman's* rich and seminal foundational work has been the useful or even indispensible role of abstract (large cardinal) set theory in lower and seemingly unrelated areas of mathematics. His paper *A Way Out* is a remarkable exercise in this methodology. A 'slight' amendment of the naïve comprehension axiom yields a single, rather strong set existence axiom, which countenances only sets and which, with no additional axioms, interprets ZFC . The price to pay for this uniformity is the appeal to a logical strength of the order of subtle cardinals, which are beyond such "small" large cardinals as Mahlo and indescribables, but still compatible with Gödel's $V = L$. — In his paper *Completeness and Iteration in Modern Set Theory*, the set theorist *S. Friedman*, who made substantial contributions to the fine structure theory, sketches a picture for the universe of sets based on two principles of completion and iteration with respect to the '# (sharp) operation', which was originally introduced as a device for transcending the constructible universe. — Finally, in this first group, the set theorist and philosopher of mathematics *K. Hauser* addresses the on-going search for "new axioms" (for additional background see [19]). In his essay *Was sind und was sollen* (*neue*) *Axiome?* he first gives an accessible overview of the technical issues involved and proceeds to interpret the principles underlying the development of modern set theory in the light of Husserl's philosophy. This intriguing project gains additional interest in view of the well-known fact that Gödel studied Husserl carefully, and because of a renewed interest in Husserl in recent general philosophy.

There has been a development in proof theory for some time to analyse weak set theories in the framework of the Kripke–Platek theory of admissible sets, KP (see [4] for the general background on KP, and [42], [59] for its proof-theoretical aspects). One of the major issues in this research program is to carry over from classical proof theory the technique of ordinal analysis, i.e., the calibration of theories in terms of their proof-theoretic strength, given by a characteristic ordinal number. The well-known limit of predicative theories is the least impredicative Feferman–Schütte ordinal Γ_0. A particular KP theory of admissible sets with this strength, KPi^0, is the point of departure for the contribution of *G. Jäger* and *D. Probst*, *Iterating* Σ *Operations in Admissible*

Set Theory Without Foundation: A Further Aspect of Metapredicative Mahlo. The system describes a recursively inaccessible universe over the natural numbers taken as urelements. If an axiom schema allowing for the iteration of Σ operations is added, a theory is arrived at whose proof-theoretic ordinal is shown to be the metapredicative[33] Mahlo ordinal $\varphi\omega 00$.[34] The result is indicative of another aspect of modern proof theory, viz., the emulation of large cardinals in a recursive setting.

The doyen of mathematical logic, *S. Feferman*, addresses the striking feature of 'typical ambiguity' in Russell's presentation of type theory that allows Russell to steer clear of the paradoxes by working in type theory and yet to avoid syntactic clutter by doing without overt type distinctions, thereby making sense, in particular, of statements like $Cls \in Cls$. Feferman has been concerned repeatedly in his own work with related issues arising in modern mathematical practice. In his paper *Typical Ambiguity: Trying to Have Your Cake and Eat It Too*, he introduces a system of reflective set-theoretic universes, conservative over ZF(C), in which a rigorous meaning can be attached to such self-reflective statements, including those of standard category theory. It is stressed that statements of the form $A \in A$ cannot be made *literally true* this way, of course; but in the author's eyes this problem of genuine self-application has yet to find a satisfactory resolution. Two approaches in the literature, theories with the anti-foundation axiom, AFA, and Quine's system of New Foundations NF, are shortly discussed in this respect.

K.-G. Niebergall is a logician and philosopher who, after Ph.D. work on the metamathematics of non-axiomatizable theories and further work on Hilbert's finitism, has been engaged for some time in an in-depth analysis of the relation of reducibility between theories, both mathematical and non-mathematical. This is borne out by his contribution *Is ZF Finitistically Reducible?* Finitistic theories in the sense of Feferman, Mycielski, and Lavine are closely scrutinized, with a special focus on the supposed paradigm of a finitistic theory, primitive recursive arithmetic, PRA. The author singles out the fact that each closed theorem of PRA has a finite model, and considers its generalization, with 'T' for 'PRA', as a candidate criterion for being a finitistic theory T. Drawing on Mycielski's locally finite theories mentioned above, the author shows that this feature has the unwelcome consequence of making ZF "finitistically reducible", certainly a problem of intuitive adequacy for reductive proof theory. — Finally, *T. Hürter's* short, but intriguing paper *Inconsistency in the Real World* constructs, in rigorous formal terms, a conceivable situation in which the conclusion is reached by contradiction that there are only finitely many natural numbers. This "worst case scenario" might be somewhat disquieting to a complacent Kroneckerian worldview.

The next three papers add the perspective of constructive mathematics to the themes of the volume. The proof-theorist *M. Rathjen*, who became known for his ground-

[33] This term is due to Jäger; it stands for the range of ordinals beyond Γ_0 whose analysis can still be carried out with methods from predicative proof theory.

[34] The function φ is a ternary version of the well-known Veblen function on the ordinals.

breaking work on Π^1_2-analysis (see [65], [66], [67]), deals here with the ban on circular definitions in the predicativist's sense, and compares it with circularity phenomena that have been studied more recently in computer science and knowledge representation by means of the anti-foundation axiom.[35] The paper *Predicativity, Circularity, and Anti-Foundation* uses a system of constructive set theory with AFA to model this kind of circularity, thereby showing that it is conceptually different from the one arising in impredicative contexts. — The logician and philosopher of mathematics *J. L. Bell*, who has been known for his extensive foundational work from the algebraic and category-theoretic viewpoint, contributes a short note entitled *Russell's Paradox and Diagonalization in a Constructive Context*. In it he points out that of the two classically equivalent methods of proving Cantor's Theorem, viz., Russell's construction and Cantor's own diagonalization method, only the former remains constructively valid, whereas the latter fails to be so. Hence, the author says, Russell's paradox has a claim to universal applicability. — The mathematician *P. Schuster* and his co-author *H. Schwichtenberg*, who has been prominently active in recent years on the interface of proof theory and computer science, choose a rigorously constructive approach to some classical problems of analysis, in particular by circumventing the axiom of countable choice. Their paper *Constructive Solutions of Continuous Equations*, while naturally technical, works with a Bishop-style notion of set (as "identified with the defining property of its elements") and hence shares its conceptualist flavor with Russell's no-classes theory.

The papers in the next group, among which H. Friedman's could also be counted, all start from the classical paradoxes and use modern tools to explore the remaining logical space in different directions. The logician and philosopher, and former Emmy Noether scholar, *K. F. Wehmeier* gives a survey of various subsystems of Frege's (second-order) logic of the *Grundgesetze* which can be shown to be consistent, among them the fragment with Δ^1_1-comprehension, which was proved consistent in joint work of the author with F. Ferreira. Apart from the question, how much mathematics could still be derived, the logicist willing to retreat to such a system would face another problem for his program, as the author points out: the system proves the existence of objects other than value-ranges. — The logician *A. Cantini*, known especially for his [8], addresses the paradox of propositions appearing in Appendix B of Russell's *Principles*. This paradox is of particular interest not only because it is genuinely Russellian, but also because it inevitably calls for an analysis of the notion of truth. In his paper *On a Russellian Paradox about Propositions and Truth* the author provides resolutions of this paradox by dealing with it in two quite different types of theories: one is an algebraic framework of propositions and truth based on combinatory logic, the other a consistent subtheory of Quine's New Foundations NF, which employs the device of stratification. — *H. Field*, whose influential work in the philosophy of mathematics was already mentioned above, investigates the option of adopting a

[35] In the obvious graph-theoretic interpretation of set-theoretic membership, this axiom allows for loops but requires a unique "decoration" for every graph.

naïve comprehension scheme for properties and building a consistent theory around it that weakens classical logic in an intricate way. This idea is developed and defended in detail in the paper *The Consistency of the Naive Theory of Properties*, where an explicit model for the resulting logic in a multi-valued semantics is constructed. — The final paper in this group, *The Significance of the Largest and Smallest Numbers for the Oldest Paradoxes*, is based on yet another and quite unique approach to the paradoxes. Its author, the logician and philosopher *U. Blau*, has been working on a comprehensive system of a multi-valued "reflexion logic" for many years,[36] pursuing Russell's global strategy of finding a uniform resolution of all known paradoxes. Another prominent feature of the approach is the distinctively Platonist and Cantorian philosophy of mathematics underlying it. The paper defines and explores a novel number system which extends the standard model of the ordinal and real numbers and sheds new light on the old mysteries of counting, dividing, and motion.

The papers in the following group were written by logician-philosophers and prominent Russell scholars. They deal with specific topics in Russell's logical philosophy. *N. Griffin*, a specialist for the early Russell, contributes the paper *The Prehistory of Russell's Paradox*, in which he gives an overview of the genesis of Russell's paradox and dwells on the question, already hinted at above, why Russell, unlike the Cantorians, was particularly prone to engage in a serious analysis of the paradoxes. — *G. Landini* has become known for his important work on Russell's substitutional theory. In his paper *Logicism's 'Insolubilia' and Their Solution by Russell's Substitutional Theory* he remarkably argues that the substitutional theory, which Russell entertained in the years 1905–1907 and which has become available from the Russell Archives in its full extant content only in recent years, contains all the ingredients not only for the resolution of both the logical and semantical paradoxes but also for successfully carrying out the logicist program. — *Ph. de Rouilhan's* paper *Substitution and Types: Russell's Intermediate Theory* is concerned with the transition period in Russell's logical philosophy between the substitutional theory and the mature theory of ramified types. Apart from two main versions of the former described in the literature the author distinguishes and gives an interpretation of a third version, which he calls "the intermediate theory" and which is contained in just a few pages of [74]. — In the next paper, *Propositional Ontology and Logical Atomism*, *F. Rodríguez-Consuegra* addresses the problems inherent in Russell's theory of logical atomism, arguing that its ontology was still haunted by the ghost of Bradley's paradox of relations, which Russell never succeeded in disposing of completely. — The next two papers deal with the logic of *Principia* proper. The first by *B. Linsky*, *Classes of Classes and Classes of Functions in 'Principia Mathematica'*, discusses a problem that was recently shown to arise in connection with the no-classes theory. It is argued that in order to avoid it a distinction is needed between classes of classes and classes of (propositional) functions. The distinction was already considered by Whitehead and Russell but not

[36] This is the system *Logik der Unbestimmtheiten und Paradoxien*, referred to in the paper, which *inter alia* subsumes the 3-valued logic in Kripke's theory of truth and various other truth theories.

pursued further; by showing how to resolve the problem with it the paper corroborates the intensional character of *Principia*'s logic even where it deals with classes. — There is the technical issue regarding a claim made by Russell in Appendix B of the second edition of *Principia*, to the effect that the amended type theory proposed there (his '1925 logic') proves unrestricted mathematical induction, with an axiom of infinity as the only non-logical assumption. The derivation given in Appendix B, however, is defective, as Gödel [28] first noticed. The 1925 logic was in fact not described in any detail until Landini's [47], where it is claimed that mathematical induction can be derived in it after all. Drawing on joint work with J. Davoren, type theory specialist *A. Hazen* provides a thorough analysis of the situation in his paper *A "Constructive" Proper Extension of Ramified Type Theory* (*The Logic of* Principia Mathematica*, Second Edition, Appendix B*). In it a semantics for the 1925 logic is defined, with a substitutional interpretation of its higher-order quantifiers. It is argued that Landini's proof, albeit correct, is based on an added extensionality axiom, which is incorrect on the given semantics and hence fails to be faithful to the sustained "constructivist" spirit of the Russell of the second edition. — The final essay in the Russell group, *Russell on Method* by *A. Irvine*, discusses two conflicting attitudes in Russell's philosophy, his foundationalism on the one hand and his fallibilism on the other. Russell's theory of knowledge is reviewed in the light of the resulting tension.

The following two papers are devoted to the historical context of the discovery of Russell's paradox. *V. Peckhaus*, who wrote a monograph on Hilbert's program and German *Kritische Philosophie*, gives an interesting account of the interaction between mathematicians and philosophers in Göttingen at the time. His paper *Paradoxes in Göttingen* discusses the Zermelo–Russell paradox and Hilbert's own paradox (see above), and describes the reaction to Russell's *Principles* and the debate on the foundations of mathematics in the circle round the philosopher Leonard Nelson. — *D. C. McCarty*, an expert on mathematical intuitionism, explores the prehistory of constructive mathematics in his contribution *David Hilbert and Paul Du Bois-Reymond: Limits and Ideals*. The paper describes the contrast between Hilbert's scientific optimism and the skeptical epistemology of the mathematician Paul du Bois-Reymond. The latter shared with his brother, the physiologist Emil, the 'Ignorabimus' Hilbert famously combated, deriving it from a constructive approach to mathematics which anticipated a number of intuitionistic themes.

The final group of papers contains reflections on various topics in the philosophy of mathematics, the philosophy of science, and the philosophy of language, all relating to foundational issues. In his paper *Russell's Paradox and Hilbert's (much Forgotten) View of Set Theory*, *J. Mycielski*, whose work on locally finite theories was alluded to above, contributes a mathematician's view to the foundational debate, which the author calls a rationalist perspective. — In the essay *Objectivity: The Justification for Extrapolation*, *S. Lavine* draws on his much acclaimed [48] to put to use the finite mathematics of the indefinitely large in the foundation of the theory of scientific measurement. — Current philosophy of mathematics distinguishes a number of struc-

turalist views on mathematical objects, their common core being that it is structures that matter in mathematics, and not the particular way in which the objects occurring in those structures are modeled. *G. Hellman*, author of "modal structuralism" [40], discusses his own and three other main varieties of structuralism in his paper *Russell's Absolutism vs.(?) Structuralism*. He relates the modern discussion with Russell's views on these matters, in which he detects both "absolutist" and structuralist elements. Finding some points in Russell's critique of Dedekind's structuralism relevant to the evaluation of its modern versions, he uses these insights to argue for adopting modal structuralism, supplemented by a category theoretic outlook. — *R. S. D. Thomas*, the editor of *Philosophia Mathematica*, argues for a 'deflationary' view regarding the importance of ontology in mathematics. In his essay *Mathematicians and Mathematical Objects* he compares the assumption of mathematical objects with "training wheels", needed at first but eventually discarded to concentrate on what he takes to be of real importance, the mathematical relations. — *H. Sturm* is an analytic philosopher with a firm background in formal logic (with special expertise, e.g., in the model theory of modal logic). His methodological reflections *Russell's Paradox and Our Conception of Properties, or: Why Semantics is No Proper Guide to the Nature of Properties* add a skeptical note to the issue of how essential self-instantiation is when it comes to developing a viable philosophical theory of properties. It is also argued that in search of such a theory it is of little use to have recourse to semantics. — In his well-known [52] *V. McGee* deals not only with formal theories of truth, but also with the phenomenon of vagueness. His intimate knowledge of this topic is reflected in the essay *The Many Lives of Ebenezer Wilkes Smith*, where modern discussions are traced back to, and evaluated in the light of, Russell's 1923 paper "Vagueness". — In yet another cross-over, finally, the formal truth theorist and specialist on interpretability logic, *A. Visser*, looks at a central issue in the philosophy of language handed down from the work of Frege and Russell. In *What Makes Expressions Meaningful? A Reflection on Contexts and Actions* he takes issue with the commonplace version of Frege's context principle according to which the sentence is the unit of meaning. It is argued that the principle is apt to block the way to a better understanding of the mechanisms of discourse.

References

[1] Abramsky, S., Dov M. Gabbay and Thomas S. E. Maibaum (eds.): 1992. *Handbook of Logic in Computer Science. Volume 2. Background: Computational Structures*. Oxford: Clarendon.

[2] Ajdukiewicz, Kazimierz: 1935. Die syntaktische Konnexität. In *Stud. Phil.* 1: 1–27. English translation in: S. McCall (ed.), *Polish Logic, 1920–1939*, Oxford 1967, 207–231.

[3] Barendregt, Henk P.: Lambda calculi with types. In [1]: 117–309.

[4] Barwise, Jon: 1975. *Admissible Sets and Structures*. Berlin: Springer.

[5] Boolos, George: 1987. The consistency of Frege's *Foundations of Arithemetic*. In: J. J. Thomson (ed.), *On Being and Saying: Essays in Honor of Richard Cartwright*, Cambridge, MA: MIT Press, 3-20. Reprinted in [11]: 211–233.

[6] Blumenthal, Otto: 1935. Lebensgeschichte. (Biographical essay of Hilbert.) In: David Hilbert, *Gesammelte Abhandlungen. Band III: Analysis. Grundlagen der Mathematik. Physik. Verschiedenes. Lebensgeschichte*, Berlin: Springer, 388–429.

[7] Buss, Samuel R. (ed.): 1998. *Handbook of Proof Theory.* Elsevier: Amsterdam.

[8] Cantini, Andrea: 1996. *Logical Frameworks for Truth and Abstraction. An Axiomatic Study.* Amsterdam: Elsevier.

[9] Church, Alonzo: 1976. Comparison of Russell's resolution of the semantical antinomies with that of Tarski. *Journal of Symbolic Logic* 41, 747–60. Reprinted (with corrections) in: R. L. Martin (ed.), *Recent Essays on Truth and the Liar Paradox*, Oxford: Clarendon 1984, 289–306.

[10] Constable, Robert L.: 1998. Types in logic, mathematics and programming. In [7]: 683–786.

[11] Demopoulos, William: 1995. *Frege's Philosophy of Mathematics.* Cambridge, MA: Harvard University Press.

[12] Feferman, Solomon: 1977. Theories of finite type related to mathematical practice. In: J. Barwise (ed.), *Handbook of Mathematical Logic*, Amsterdam: North Holland, 913–971.

[13] Feferman, Solomon: 1988. Hilbert's program relativized: Proof-theoretical and foundational reduction. *Journal of Symbolic Logic* 53: 364–384.

[14] Feferman, Solomon: 1988. Weyl vindicated: "Das Kontinuum" 70 years later. In: *Temi e prospettive della logica e della filosofia della scienza contemporanee*. Vol. I. Bologna: CLUEB, 59–93. Reprinted in [16]: chap. 13, 249–283.

[15] Feferman, Solomon: 1993. Why a little bit goes a long way: Logical foundations of scientifically applicable mathematics. In: *PSA 1992*. Vol. 2. East Lansing: Philosophy of Science Association, 442–455. Reprinted in [16]: chap. 14, 284–298.

[16] Feferman, Solomon: 1998. *In the Light of Logic.* New York, Oxford: Oxford University Press.

[17] Feferman, Solomon, and Gerhard Jäger: 1993. Systems of explicit mathematics with non-constructive μ-operator, I. *Annals of Pure and Applied Logic* 65: 243–263.

[18] Feferman, Solomon, and Gerhard Jäger: 1993. Systems of explicit mathematics with non-constructive μ-operator, II. *Annals of Pure and Applied Logic* 79: 37–52.

[19] Feferman, Solomon, Harvey M. Friedman, Penelope Maddy, and John R. Steel: 2000. Does mathematics need new axioms? *Bulletin of Symbolic Logic* 6 (4): 401–446.

[20] Feyerabend, Paul: 1978. *Against Method*. London: Verso.

[21] Field, Hartry: 1980. *Science without Numbers. A Defence of Nominalism*. Princeton: Princeton University Press.

[22] Fine, Kit: 1985. *Reasoning with Arbitrary Objects*. Oxford: Basil Blackwell.

[23] Frege, Gottlob: 1883. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*. Vol. I. Jena: Pohle.

[24] Gabriel, Gottfried, Friedrich Kambartel and Christian Thiel (eds.): 1980. *Gottlob Freges Briefwechsel mit D. Hilbert, E. Husserl, B. Russell, sowie ausgewählte Einzelbriefe Freges*. Hamburg: Meiner.

[25] Galilei, Galileo: 1623. *Il Saggiatore*. English: *The Assayer*. Quoted from the abridged translation in: *Discoveries and Opinions of Galileo. Translated with an Introduction and Notes by Stillman Drake*, Garden City, NY: Doubleday 1957.

[26] Gallin, Daniel: 1975. *Intensional and Higher-Order Modal Logic. With Applications to Montague Semantics*. Amsterdam: North-Holland.

[27] Gödel, Kurt: 1931. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38, 173–198. Reprinted, with an English translation, in [31]: 144–194.

[28] Gödel, Kurt: 1944. Russell's mathematical logic. In: P. A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, LaSalle: Open Court, 125–153. Reprinted in [58], 192–226, and in [32]: 119–141.

[29] Gödel, Kurt: 1947. What is Cantor's continuum problem? *American Mathematical Monthly* 54, 515–525. Reprinted in [32]: 176–187.

[30] Gödel, Kurt: 1958. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunkts. *Dialectica* 12: 280–287. Reprinted, with an English translation, in [32]: 240–251.

[31] Gödel, Kurt: 1986. *Collected Works. Volume I: Publications 1929–1936*. Edited by S. Feferman *et al*. Oxford: Oxford University Press.

[32] Gödel, Kurt: 1990. *Collected Works. Volume II: Publications 1938–1974*. Edited by S. Feferman *et al*. Oxford: Oxford University Press.

[33] Gödel, Kurt: 1995. *Collected Works. Volume III: Unpublished Essays and Lectures*. Edited by S. Feferman *et al*. Oxford: Oxford University Press.

[34] Gödel, Kurt: 2003. *Collected Works. Volume V: Correspondence H–Z*. Edited by S. Feferman and J. W. Dawson Jr. *et al*. Oxford: Oxford University Press.

[35] Goldfarb, Warren D.: 1979. Logic in the twenties: The nature of the quantifier. *Journal of Symbolic Logic* 44: 351–68.

[36] Goodman, Nelson and Willard V. Quine: 1947. Steps toward a constructive nominalism. *Journal of Symbolic Logic* 12: 105–122.

[37] Griffin, Nicholas: 1991. *Russell's Idealist Apprenticeship*. Oxford: Clarendon.

[38] Hazen, Allen P.: 1983. Predicative logics. In: D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic. Volume I: Elements of Classical Logic*, Dordrecht: Reidel, 331–407.

[39] Heck, Richard G. Jr.: 1993. The development of arithmetic in Frege's *Grundgesetze der Arithmetik*. *Journal of Symbolic Logic* 58: 579–601. Reprinted with minor revisions and a Postscript in [11]: 257–294.

[40] Hellman, Geoffrey: 1989. *Mathematics withour Numbers. Towards a Modal-Structural Interpretation*. Oxford: Clarendon.

[41] Hylton, Peter: 1990. *Russell, Idealism, and the Emergence of Analytic Philosophy*. Oxford: Clarendon.

[42] Jäger, Gerhard: 1986. *Theories for Admissible Sets. A Unifying Approach to Proof Theory*. Napoli: Bibliopolis.

[43] Kanamori, Akihiro: 1997. *The Higher Infinite. Large Cardinals in Set Theory from Their Beginnings*. Berlin: Springer.

[44] Kant, Immanuel: 1787. *Kritik der reinen Vernunft*. 2nd edition. Hartknoch: Riga. English translation: *Critique of Pure Reason*. London: Macmillan 1933.

[45] Klein, Felix: 1926–27. *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert. Teil I, II*. Berlin: Springer.

[46] Kuhn, Thomas S.: 1978. *Black-Body Theory and the Quantum Discontinuity, 1894–1912*. Chicago: University of Chicago Press.

[47] Landini, Gregory: 1996. The definability of the set of natural numbers in the 1925 *Principia Mathematica*. *Journal of Philosophical Logic* 25: 597–615.

[48] Lavine, Shaughan: 1994. *Understanding the Infinite*. Cambridge, MA: Harvard University Press.

[49] Link, Godehard: 2000. Reductionism as resource-conscious reasoning. *Erkenntnis* 53: 173–193.

[50] Mancosu, Paolo: 2003. The Russellian influence on Hilbert and his school. *Synthese* 137: 59–101.

[51] Martin-Löf, Per: *Intuitionistic Type Theory*. Napoli: Bibliopolis.

[52] McGee, Vann: 1991. *Truth, Vagueness, and Paradox. An Essay on the Logic of Truth*. Indianapolis: Hackett.

[53] Montague, Richard: 1974. *Formal Philosophy*. Edited and with an introduction by Richmond H. Thomason. New Haven: Yale University Press.

[54] Moore, Gregory H.: 1988. The roots of Russell's paradox. *Russell. The Journal of the Bertrand Russell Archives* 8: 46–56.

[55] Moortgat, Michael: 1997. Categorial type logics. In: J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Amsterdam: Elsevier, 93–177.

[56] Mycielski, Jan: 1986. Locally finite theories. *Journal of Symbolic Logic* 51: 59–62.

[57] Peckhaus, Volker, and Reinhard Kahle: 2002. Hilbert's paradox. *Historia Mathematica* 29: 157–175.

[58] Pears, David S. (ed.): 1972. *Bertrand Russell. A Collection of Critical Essays*. Garden City, NY: Doubleday.

[59] Pohlers, Wolfram: 1998. Subsystems of set theory and second order number theory. In [7]: 209–335.

[60] Putnam, Hilary: 1999. *The Threefold Cord. Mind, Body, and World*. New York: Columbia University Press.

[61] Quine, Willard V.: 1960. *Word and Object*. Cambridge, MA: MIT Press.

[62] Quine, Willard V.: 1972. Remarks for a memorial symposium. In [58]: 1–5.

[63] Quine, Willard V.: 1986. Reply to Charles Parsons. In: L. E. Hahn and P. A. Schilpp (eds.), *The Philosophy of W. V. Quine*, LaSalle: Open Court, 396–403.

[64] Quine, Willard V.: 1991. Immanence and validity. *Dialectica* 45: 219–230.

[65] Rathjen, Michael: 1995. Recent advances in ordinal analysis: Π^1_2-CA and related systems. *Bulletin of Symbolic Logic* 4: 468–485.

[66] Rathjen, Michael: To appear. An ordinal analysis of stability. *Archive for Mathematical Logic*.

[67] Rathjen, Michael: To appear. An ordinal analysis of parameter-free Π^1_2 comprehension. *Archive for Mathematical Logic*.

[68] Reid, Constance: 1970. *Hilbert. With an Appreciation of Hilbert's Mathematical Work by Hermann Weyl*. Berlin: Springer.

[69] Robinson, Abraham: 1966. *Non-Standard Analysis*. Amsterdam: North-Holland.

[70] Russell, Bertrand: 1901. Part I of the *Principles*, Draft of 1901. In: *The Collected Papers of Bertrand Russell*, vol. 3: *Towards the "Principles of Mathematics" 1900–1902*, London: Routledge, 1993, 181–208.

[71] Russell, Bertrand: 1903. *The Principles of Mathematics*. London: Allen & Unwin.

[72] Russell, Bertrand: 1905. On denoting. *Mind* 14, 479–493. Reprinted in [79]: 41–56.

[73] Russell, Bertrand: 1907. The regressive method of discovering the premises of mathematics. In [81]: 272–283.

[74] Russell, Bertrand: 1908. Mathematical logic as based on the theory of types. *American Journal of Mathematics* 30: 222–262. Reprinted in [79]: 59–102, and in [84]: 150–182.

[75] Russell, Bertrand: 1913. *Theory of Knowledge. The 1913 Manuscript*. Edited by E. R. Eames in collaboration with K. Blackwell. London: Routledge 1992.

[76] Russell, Bertrand: 1914. The relation of sense-data to physics. *Scientia* 14: 1–27. Reprinted in [77]: chap. 8.

[77] Russell, Bertrand: 1917. *Mysticism and Logic*. London: Allen & Unwin.

[78] Russell, Bertrand: 1919. *Introduction to Mathematical Philosophy*. London: Unwin. Reprinted by Routledge.

[79] Russell, Bertrand: 1956. *Logic and Knowledge. Essays 1901–1950*. Edited by R. C. Marsh, London: Unwin.

[80] Russell, Bertrand: 1959. *My Philosophical Development*. London: Allen & Unwin. Reprinted by Routledge.

[81] Russell, Bertrand: 1973. *Essays in Analysis*. London: Allen & Unwin.

[82] Simpson, Stephen G.: 1999. *Subsystems of Second Order Arithmetic*. Berlin: Springer.

[83] Solovay, Robert M.: 1995. Introductory note to *Gödel *1939b* and *Gödel *1940a*. In [33]: 114–126.

[84] van Heijenoort, Jean (ed.): 1967. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press.

[85] Whitehead, Alfred N. and Bertrand Russell: 1910–13. *Principia Mathematica*. Cambridge: Cambridge University Press. Second edition, 1925–27.

[86] Woodin, W. Hugh: 1994. Large cardinal axioms and independence: The continuum problem revisited. *The Mathematical Intelligencer* 16 (no. 3): 31–36.

[87] Woodin, W. Hugh: 1999. *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*. Berlin: De Gruyter.

[88] Woodin, W. Hugh: 2004. Set theory after Russell: The journey back to Eden. In: G. Link (ed.), *One Hundred Years of Russell's Paradox. Mathematics, Logic, Philosophy.* Berlin: Walter de Gruyter 2004, 29–47.

Seminar für Philosophie, Logik and Wissenschaftstheorie
Philosophie-Department
Universität München
Ludwigstraße 31/I
80539 München
Germany

E-mail: glink@lrz.uni-muenchen.de

Set Theory after Russell: The Journey Back to Eden

W. Hugh Woodin

Abstract. Can the fundamental questions such as the *Continuum Hypothesis* be solved? Or are these questions simply questions without answers? Are there any questions in Set Theory which are formally unsolvable and yet answerable? If so what distinguishes these questions from the *Continuum Hypothesis* and is this distinction a meaningful one?

The issues are examined in the context of Ω-logic.

1. Introduction

Does the phenomenon of formal independence in Set Theory fulfill the prophecy some might claim is the content of Russell's discovery of the now famous *Russell Paradox*? This claim of course is that there can be no meaningful axiomatization of Set Theory because the concept of set is inherently vague, moreover any choice of axioms is an arbitrary one and this is the reason that to date essentially all investigations have rather quickly led to inconsistency or to unsolvable problems.

Perhaps the most well known example of a problem of Set Theory which is known to be formally unsolvable is the problem of Cantor's *Continuum Hypothesis*. Does the formal independence of the *Continuum Hypothesis* from the axioms of Set Theory imply that the question has no meaning? Some, including Cohen himself, have argued that the answer to this question is "yes" [1]. But for this position to be credible either the other classical problems in set theory which have been similarly shown to be formally unsolvable would also have to be regarded as without meaning, or there must be an explanation as to why the problem of the *Continuum Hypothesis* is different. I have in mind here the classical questions of descriptive set theory for they arose not long after the *Continuum Hypothesis* was first posed.[1]

The *projective sets* are those subsets of $\mathbb{R}$ which can be generated from the closed sets in finitely many steps of taking continuous images, by continuous functions

$$f : \mathbb{R} \to \mathbb{R},$$

[1] Some of the issues to be discussed here have been discussed in two articles which have appeared in the *Notices of the AMS*, [12] and [13] but there is an important expository change in this account which concerns the definition of Ω-logic. There are also some new results mentioned in this account.

or taking complements. One classical question is simply: Are the projective sets Lebesgue measurable?

The sets which can be generated in 4 steps are the analytic sets and their complements. It is a classical theorem that these sets are Lebesgue measurable. What about the sets generated in 5 steps? This question is formally unsolvable. So is this question without meaning? If not, then how is the problem of the *Continuum Hypothesis* any different?

2. The Structure of All Sets of Finite Ordinals

The structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle$$

is of course the standard structure for *Second-Order Number Theory*.

Within this structure one can formulate many of the classical questions concerning the projective sets, including the question on the Lebesgue measurability of the projective sets generated in 5 steps from the closed sets. Therefore if the axioms for set theory are not sufficient to formally provide answers, one must seek new axioms.

A set $A \subseteq \mathcal{P}(\omega)$ is *definable* in the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle,$$

from parameters if there exists a formula $\phi(x, y)$ in the first order language for this structure and $b \in \mathcal{P}(\omega)$ such that

$$A = \{a \in \mathcal{P}(\omega) \mid \langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle \vDash \phi[a, b]\}.$$

The projective sets correspond with subsets $A \subseteq \mathcal{P}(\omega)$ such that A can be defined, from parameters, in the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle.$$

More precisely, let $\pi : \mathcal{P}(\omega) \to [0, 1]$ be the surjection where for each $a \subseteq \omega$,

$$\pi(a) = \sum_{i \in a} 2^{-i}$$

if $a \neq \emptyset$ and $\pi(\emptyset) = 0$. Then $X \subseteq [0, 1]$ is a projective set if and only if the preimage of X under π is definable from parameters in the structure, $\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle$.

Now suppose $A \subseteq \mathcal{P}(\omega)$. Associated to the set A is an infinite game, G_A, involving two players. The players alternate choosing elements of $\{0, 1\}$ with Player I moving first. After ω many moves a sequence $\langle \epsilon_i : i < \omega \rangle$ is specified. Let

$$a = \{i < \omega \mid \epsilon_i = 1\}.$$

Player I wins this run of the game if $a \in A$. Otherwise Player II wins. When a player moves in this game the state of the game is a finite sequence of 0's and 1's; the sequence is of even length when it is Player I's turn to move, and the sequence is of odd length when it is Player II's turn to move. A strategy is a function,

$$\tau : \text{SEQ} \to \{0, 1\}$$

where SEQ is the set of all finite sequences of 0's and 1's. The strategy τ is a winning strategy for Player I if following the strategy Player I always wins, no matter how Player II plays. Similarly the strategy τ is a winning strategy for Player II if following the strategy Player II always wins, no matter how Player I plays. The game, G_A, is *determined* if there is a winning strategy for one of the players. Clearly, there can be a winning strategy for at most one of the two players. It is a consequence of the *Axiom of Choice* that there exist sets $A \subset \mathcal{P}(\omega)$ for which neither player has a winning strategy in the associated game, G_A. So the axiom which asserts that all of the games G_A are determined is inconsistent (with the axioms of set theory).

Projective Determinacy is the axiom:

> *For each set $A \subseteq \mathcal{P}(\omega)$ such that A is definable from parameters in the structure,*
>
> $$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle,$$
>
> *the game G_A is determined.*

This axiom is the correct (and missing) axiom for the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle.$$

Projective Determinacy settles (in the context of ZFC) the classical questions concerning the projective sets and moreover Cohen's method of forcing *cannot* be used to establish that questions of second order number theory are formally unsolvable from this axiom.

Finally the intricate connections between this axiom and large cardinal axioms provide compelling evidence that this axiom is *true*. For me, granting the truth of the axioms for Set Theory, the *only* conceivable argument against the truth of this axiom, would be its inconsistency. I also claim that, at present, the only credible basis for the belief that the axiom is consistent is the belief that the axiom is true. This state of affairs could change as the number theoretic consequences of the axiom become more fully understood.

It is interesting to note that the understanding of this axiom took quite a number of years of research and that the credible claim for the truth of the axiom was only possible after this research was accomplished—there is no known elementary argument for the truth of the axiom. This fact creates something of a challenge for the skeptic to explain.

Should the axiom of *Projective Determinacy* be accepted as true? One argument that is frequently cited against this axiom is that there are interesting alternative theories for the projective sets, indeed there are some very difficult open questions about the

possible theories for the projective sets in the context where *Projective Determinacy* is false. But this argument confuses two issues. Accepting *Projective Determinacy* as true does not deny the study of models in which it is false. For example, it is certainly the prevailing belief that the axioms for Number Theory are *consistent*. Nevertheless the study of models of Number Theory in which the axioms are *inconsistent* seems clearly an interesting program and a deep understanding of these models probably entails resolving the P = NP question of computational complexity.

Do I believe the axiom of *Projective Determinacy* is true? This I find to be a rather frequently asked question of me. My answer is this. I believe the axiom of *Projective Determinacy* is as true as the axioms of Number Theory. So I suppose I advocate a position that might best be described as *Conditional Platonism*.

3. The Structure of All Sets of Countable Ordinals

What is the next structure, from the standpoint of complexity, to consider beyond the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle?$$

Letting $\mathcal{P}(\mathrm{Ord})$ denote the class of all sets of ordinals, it is a consequence of the *Axiom of Choice* that the universe of sets is logically equivalent to the structure,

$$\langle \mathcal{P}(\mathrm{Ord}), \mathrm{Ord}, \cdot, +, \in \rangle.$$

Therefore a natural hierarchy of structures is given by the sequence of structures,

$$\langle \mathcal{P}(\alpha), \alpha, \cdot, +, \in \rangle,$$

where α is an (infinite) ordinal which is closed under the operation $\cdot$ (so for all $\beta < \alpha$ and for all $\gamma < \alpha$, $\beta \cdot \gamma < \alpha$).

For each such countable ordinal, α, the structure

$$\langle \mathcal{P}(\alpha), \alpha, \cdot, +, \in \rangle$$

is *reducible* to the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle.$$

Therefore I claim that the next natural structure is

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

which is the standard structure for all sets of countable ordinals. It is well known that this structure is equivalent to the structure of all sets of hereditary cardinality less than $\aleph_2$.

Whether or not CH holds is a first order property of this structure. Therefore any reasonably complete axiomatization for this structure should resolve CH.

As I have noted, *Projective Determinacy* is the correct axiomatization for the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle.$$

Is there a generalization of this axiom to the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle?$$

This of course is a vague question. To make it more precise requires a logic which transcends first order logic. Before discussing the relevant logic and to help motivate the definition, I note the following theorem which gives an interesting reformulation for an axiom slightly stronger than *Projective Determinacy*.

$L(\mathbb{R})$ is Gödel's constructible universe relativized to $\mathbb{R}$; it is the smallest inner model of the universe containing $\mathbb{R}$, the ordinals, and in which the axioms of Set Theory hold (except possibly the Axiom of Choice). AD is the axiom that for all sets $A \subseteq \mathcal{P}(\omega)$, the game G_A is determined. As I have previously noted, this axiom contradicts the Axiom of Choice. However the axiom

$$L(\mathbb{R}) \vDash \text{AD}$$

is not obviously inconsistent since the Axiom of Choice can fail to hold in $L(\mathbb{R})$; if $\mathbb{B}$ is the complete Boolean algebra given by the partial order that Cohen originally used to establish the consistency of the negation of the *Continuum Hypothesis*, then

$$V^{\mathbb{B}} \vDash \text{“}L(\mathbb{R}) \nvDash \text{Axiom of Choice”}.$$

In fact the axiom, “$L(\mathbb{R}) \vDash$ AD”, is a natural strengthening of *Projective Determinacy* since

$$\mathcal{P}(\mathbb{R}) \cap L(\mathbb{R})$$

is a natural extension of the projective sets.

If there exists a proper class of Woodin cardinals[2] then

$$L(\mathbb{R}) \vDash \text{AD};$$

and moreover for all complete Boolean algebras, $\mathbb{B}$,

$$V^{\mathbb{B}} \vDash \text{“}L(\mathbb{R}) \vDash \text{AD”}.$$

Theorem 1. *Suppose there exists a proper class of strongly inaccessible cardinals. The following are equivalent:*

1. *For all complete Boolean algebras,* $\mathbb{B}$*,*

$$V^{\mathbb{B}} \vDash \text{“}\mathcal{P}(\omega_1) \notin L(\mathbb{R})\text{”}.$$

[2]For the definition of a Woodin cardinal and related historical remarks, see [6].

2. *For all complete Boolean algebras,* $\mathbb{B}$,

$$V^{\mathbb{B}} \models \text{“}L(\mathbb{R}) \models \text{AD”}.$$

3. *For all complete Boolean algebras,* $\mathbb{B}$,

$$V^{\mathbb{B}} \models \text{“}L(\mathbb{R}) \not\models \text{Axiom of Choice”}. \qquad \square$$

Thus if there exists a proper class of strongly inaccessible cardinals and if the theory of the inner model $L(\mathbb{R})$ is generically absolute then necessarily, $L(\mathbb{R}) \models \text{AD}$. Finally this is a nontrivial claim; if there is a proper class of strongly inaccessible cardinals then in Gödel's inner model, L, there is a proper class of strongly inaccessible cardinals but in L the axiom of *Projective Determinacy* is false. In other words, the assumption that there is a proper class of strongly inaccessible cardinals does not imply $L(\mathbb{R}) \models \text{AD}$.

4. Ω-Logic

I define Ω-logic. Following the exposition of [2], I first define when a sentence ϕ in the language for Set Theory is Ω-valid as a consequence of a theory T (and eliminate the definition of Ω^*-logic).

Suppose that T is a set of sentences in the language for Set Theory and that ϕ is a sentence. Then $T \models_\Omega \phi$ if for all complete Boolean algebras, $\mathbb{B}$, for all ordinals α, if

$$V_\alpha^{\mathbb{B}} \models T$$

then $V_\alpha^{\mathbb{B}} \models \phi$.

A remarkable fact is that if there is a proper class of Woodin cardinals then for each theory, T, and for each sentence ϕ, the relation, $T \models_\Omega \phi$, is generically absolute. It is this fact that makes the notion, $T \models_\Omega \phi$, interesting.

Theorem 2 (ZFC). *Suppose that there is a proper class of Woodin cardinals,* T *is a set of sentences and that* ϕ *is a sentence. Then for each complete Boolean algebra* $\mathbb{B}$,

$$V \models \text{“}T \models_\Omega \phi\text{”}$$

if and only if

$$V^{\mathbb{B}} \models \text{“}T \models_\Omega \phi\text{”}. \qquad \square$$

The nontrivial direction is showing that if

$$V^{\mathbb{B}} \models \text{“}T \models_\Omega \phi\text{”}$$

then $V \models \text{“}T \models_\Omega \phi\text{”}$.

The notion that $T \vdash_\Omega \phi$ is more complicated to define and involves the definition of a transfinite hierarchy of subsets of $\mathbb{R}$ which extends the hierarchy of the projective subsets of $\mathbb{R}$. I caution that the definition given here differs slightly from the definitions given in previous accounts. The change is motivated by the definition of $T \vDash_\Omega \phi$ which eliminates the need for defining Ω^*-logic[3]. Thus the emphasis is now placed on the satisfaction relation, $T \vDash_\Omega \phi$, and the definition of the proof relation, $T \vdash_\Omega \phi$, becomes simply an attempt to isolate the corresponding notion of proof. Finally the definition given here of the relation, $T \vdash_\Omega \phi$, is given just in the context of ZFC, no additional set theoretic assumptions are made.

A set of reals, $A \subseteq \mathbb{R}$, is *universally Baire* if for every compact Hausdorff space, Ω, and for every continuous function

$$F : \Omega \to \mathbb{R},$$

the set $\{x \in \Omega \mid F(x) \in A\}$ has the property of Baire: i.e., there exists an open set $O \subseteq \Omega$ such that the symmetric difference,

$$\{x \in \Omega \mid F(x) \in A\} \triangle O,$$

is meager, [4].

Assuming the existence of a proper class of Woodin cardinals then every projective set is universally Baire. The following theorem shows that a much stronger claim is true.

Theorem 3. *Suppose that there exists a proper class of Woodin cardinals and that $A \subseteq \mathbb{R}$ is universally Baire. Then*

1. $L(A, \mathbb{R}) \vDash \mathrm{AD}^+$,
2. *every set $B \in \mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R})$ is universally Baire.* □

Here the axiom, AD^+, is a technical variation of the axiom, AD, with the feature that the assertion that $L(A, \mathbb{R}) \vDash \mathrm{AD}^+$ is arguably the correct generalization of the assertion that $L(\mathbb{R}) \vDash \mathrm{AD}$, [11]. In fact if $L(\mathbb{R}) \vDash \mathrm{AD}$ then necessarily, $L(\mathbb{R}) \vDash \mathrm{AD}^+$ but this is not an easy thing to prove.

Suppose that $A \subseteq \mathbb{R}$ is universally Baire and that M is a countable transitive model of ZFC. The set M is A-closed if for all countable transitive models, N, such that N is a set generic extension of M,

$$A \cap N \in N.$$

The notion that M is A-closed can be defined without forcing [13].

Suppose that T is a theory and that ϕ is a sentence. Then $T \vdash_\Omega \phi$ if there exists a set $A \subseteq \mathbb{R}$ such that

1. $L(A, \mathbb{R}) \vDash \mathrm{AD}^+$,

[3] In fact, the relation, $T \vDash_\Omega \phi$, is the relation, $T \vdash_{\Omega^*} \phi$, of the previous accounts.

2. every set in $\mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R})$ is universally Baire,

3. for all countable transitive A-closed sets M, for all ordinals $\alpha \in M$, if $M_\alpha \vDash T$ then $M_\alpha \vDash \phi$.

The universally Baire set, A, is the "Ω-proof". There is a natural notion of the length of this proof which is given by the ordinal rank of A in the Wadge hierarchy of complexity [13]. This in turn allows one to define the notion of a shortest Ω-proof of ϕ from T. The definition works by virtue of the following theorem.

Theorem 4 ((ZFC)). *Suppose that $A \subseteq \mathbb{R}$, $B \subset \mathbb{R}$ and that A and B each witness that $T \vdash_\Omega \phi$. Then there exists a set*

$$C \in \mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R}) \cap L(B, \mathbb{R})$$

such that C witnesses $T \vdash_\Omega \phi$. □

With this definition of length, one can define the usual sorts of Gödel sentences etc. Thus in many respects Ω-logic is a natural transfinite generalization of first order logic.

Generally for the analysis of $T \vDash_\Omega \phi$ one assumes the existence of a proper class of Woodin cardinals and for the analysis of $T \vdash_\Omega \phi$, one assumes that for all universally Baire sets A,

$$L(A, \mathbb{R}) \vDash \mathrm{AD}^+,$$

and that every set in $\mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R})$ is universally Baire–both of these assertions necessarily hold assuming there is a proper class of Woodin cardinals. Nevertheless it is convenient to define the relations, $T \vDash_\Omega \phi$ and $T \vdash_\Omega \phi$, without all of these additional assumptions. However note that the definition of $T \vdash_\Omega \phi$ is vacuous if interpreted in any inner model of the form $L[A]$ where A is a set.

Theorem 5 (Ω-soundness (ZFC)). *Suppose that T is a set of sentences, that ϕ is a sentence, and that $T \vdash_\Omega \phi$. Then $T \vDash_\Omega \phi$.* □

A natural question is why this definition of $T \vdash_\Omega \phi$ should be the correct one. A brief justification is as follows. First for each $a \in \mathbb{R}$, for each set of sentences T, and for each formula $\phi(x_0)$ one can generalize the previous definition in a natural fashion and define the relation $T \vDash_\Omega \phi[a]$. More precisely $T \vDash_\Omega \phi[a]$ if for all complete Boolean algebras, $\mathbb{B}$, for all ordinals α, if

$$V_\alpha^{\mathbb{B}} \vDash T$$

then $V_\alpha^{\mathbb{B}} \vDash \phi[a]$.

Define a set $A \subset \mathbb{R}$ to be Ω-*finite* if there exists a formula $\phi(x_0)$ such that

$$A = \{a \in \mathbb{R} \mid \emptyset \vDash_\Omega \phi[a]\}$$

and such that for all complete Boolean algebras, $\mathbb{B}$, the following holds in $V^{\mathbb{B}}$:

For all $a \in \mathbb{R}$ either $\emptyset \vDash_\Omega \phi[a]$ or $\emptyset \vDash_\Omega (\neg\phi)[a]$.

Now suppose that $\emptyset \vDash_\Omega \phi$. Then by analogy with first order logic, there should exist a set $A \subseteq \mathbb{R}$ such that A is Ω-finite and such that for all countable transitive models, M, of ZFC, if M is suitably closed under A then

$$M \vDash \text{“}\emptyset \vDash_\Omega \phi\text{”}.$$

Given the definition of the relation, $\emptyset \vDash_\Omega \phi$, the requirement that M be suitably closed under A should be:

> For all Boolean algebras, $\mathbb{B}_M \in M$, if $\mathbb{B}$ is the completion of $\mathbb{B}_M$ and if $G \subseteq \mathbb{B}$ is V-generic then
>
> $$A_G \cap M[G \cap \mathbb{B}_M] \in M[G \cap \mathbb{B}_M]$$
>
> where A_G is the set of $a \in \mathbb{R}^{V[G]}$ such that
>
> $$V[G] \vDash \text{“}\emptyset \vDash_\Omega \phi_A[a]\text{”},$$
>
> and where $\phi_A(x_0)$ is a formula witnessing that A is Ω-finite.

In short, the Ω-finite sets should suffice to “witness” the relation $\emptyset \vDash_\Omega \phi$, at least for countable transitive models.

Theorem 6 (ZFC). *Suppose that there is a proper class of Woodin cardinals and that $A \subseteq \mathbb{R}$ is Ω-finite. Then every set in $\mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R})$ is universally Baire and $L(A, \mathbb{R}) \vDash \mathrm{AD}^+$.* □

Now suppose that there is a proper class of Woodin cardinals. It follows that for each Ω-finite set A, if M is a countable transitive set such that

$$M \vDash \mathrm{ZFC}$$

and such that M is A-closed then M is suitably closed under A in the sense defined above. This suggests the definition of the relation, $T \vdash_\Omega \phi$, and the Soundness Theorem shows that the definition is not too strong.

The Ω *Conjecture* asserts that if there is a proper class of Woodin cardinals then for each sentence ϕ, if $\emptyset \vDash_\Omega \phi$ then $\emptyset \vdash_\Omega \phi$. Phrased this way the Ω Conjecture is simply a generalization of the Gödel Completeness Theorem. It is important that the Ω Conjecture be conditioned on the existence of large cardinals which are not significantly weaker than Woodin cardinals. This claim is evident from the following theorem.

Theorem 7 (ZFC). *Suppose there is a proper class of inaccessible limits of Woodin cardinals. Then there exists a transitive inner model N such that*

$N \vDash \mathrm{ZFC} +$ “ There is a proper class of inaccessible limits of Woodin cardinals”

and such that for all $\kappa < \delta$, where δ is the least Woodin cardinal of N, if κ is strongly inaccessible in N then in N_κ the following hold:

1. $\emptyset \vDash_\Omega$ "There are no Woodin cardinals"*;*
2. $\emptyset \nvDash_\Omega$ "There are no Woodin cardinals"*;*
3. *For every universally Baire set,* $A \subseteq \mathbb{R}$*,* $L(A, \mathbb{R}) \vDash \mathrm{AD}^+$ *and every set in* $\mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R})$ *is universally Baire.* □

By varying the choice of κ one can arrange that in N_κ there is a proper class of cardinals satisfying any given large cardinal notion "strictly weaker" than that of being a Woodin cardinal. One difficulty (but this is not the main difficulty) in producing the inner model N of the theorem lies in controlling the universally Baire sets of N_κ where κ is below the least Woodin cardinal of N; i.e., in an inner model with no Woodin cardinals. For example one can show that in L, every set of reals is the continuous image of a universally Baire set. This generalizes to the rank initial segments of the inner model of one Woodin cardinal below the Woodin cardinal and to the rank initial segments of all suitably closed fine structural inner models[4] below the least Woodin cardinal of the model. Another subtle aspect is that this property of these rank initial segments (that every set of reals is the continuous image of a universally Baire set) also holds of their set generic extensions for which no reals are added. So in these rank initial segments forcing notions which are ω-closed can add universally Baire sets, something which cannot happen if there is a Woodin cardinal.

The formally stronger conjecture that if there is a proper class of Woodin cardinals then for all theories, T, and for all sentences, ϕ, if $T \vDash_\Omega \phi$ then $T \vdash_\Omega \phi$, is probably equivalent to the Ω Conjecture. At least the current approaches to proving the Ω Conjecture would, if successful, prove this stronger conjecture as well. Both conjectures are consistent. For example, if the theory,

$$\mathrm{ZFC} + \text{"There is a proper class of Woodin cardinals"},$$

is consistent then so is the theory,

$$\mathrm{ZFC} + \text{"There is a proper class of Woodin cardinals"} + \text{"The } \Omega \text{ Conjecture holds"}.$$

This is also evidence that the definition of $T \vdash_\Omega \phi$ is correct. It is not known if the Ω Conjecture is consistently false.

For each sentence ϕ, ϕ is Ω_{ZFC}-valid if $\mathrm{ZFC} \vDash_\Omega \phi$ and ϕ is Ω_{ZFC}-provable if $\mathrm{ZFC} \vdash_\Omega \phi$.

Assume there is a proper class of Woodin cardinals. Then for each sentence ϕ in the language for the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle,$$

either the sentence "$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle \vDash \phi$" is Ω_{ZFC}-provable or the sentence

$$\text{"}\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle \vDash (\neg\phi)\text{"}$$

is Ω_{ZFC}-provable.

[4] Inner models of the form $L[E]$ as defined by Mitchell-Steel in [8].

This of course must fail for the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

and that is the basic problem. However there is a variation which is not obviously ruled out. Before stating the question it is convenient to fix some more notation. Suppose ψ and ϕ are sentences. Then ϕ is Ω_{ZFC}-provable from ψ if the implication

$$(\psi \to \phi)$$

is Ω_{ZFC}-provable. Thus ϕ is Ω_{ZFC}-provable from ψ if $\text{ZFC} \cup \{\psi\} \vdash_\Omega \phi$. A sentence ψ is Ω_{ZFC}-consistent if its negation $(\neg\psi)$ is not Ω_{ZFC}-provable.

The search for an absoluteness theorem for the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

leads naturally to the following question.

> *Suppose there is a proper class of Woodin cardinals. Is there an Ω_{ZFC}-consistent sentence Ψ such that for all sentences ϕ in the language for the structure,*
>
> $$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$
>
> *either the sentence "$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi$" is Ω_{ZFC}-provable from Ψ or the sentence "$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash (\neg\phi)$" is Ω_{ZFC}-provable from Ψ?*

The answer is "yes" and I shall discuss one example of such an axiom shortly. For the sake of brevity, I say an axiom, Ψ, is *good*; i.e., good for the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

if Ψ is both Ω_{ZFC}-consistent and has the property that for all sentences ϕ in the language for the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

either the sentence "$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi$" is Ω_{ZFC}-provable from Ψ or the sentence

$$\text{"}\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle \vDash (\neg\phi)\text{"}$$

is Ω_{ZFC}-provable from Ψ.

The relevance of good sentences to the problem of CH is given in the following theorem.

Theorem 8 (ZFC). *Suppose that there is a proper class of Woodin cardinals and suppose that Ψ is a sentence which is good. Then*

$$\text{ZFC} + \Psi \vdash_\Omega (\neg\text{CH}). \qquad \square$$

In other words, if the theory of the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

is to be resolved on the basis of a good axiom then necessarily CH is false. Since there are good axioms, one has an argument that CH is false based not on a specific choice of an axiom but rather based simply on a completeness property the axiom is required to have.

A sentence Ψ is Ω_{ZFC}-satisfiable if its negation, $(\neg\Psi)$, is not Ω_{ZFC}-valid. Thus Ψ is Ω_{ZFC}-satisfiable if there exist a complete Boolean algebra and an ordinal α such that

$$V_\alpha^{\mathbb{B}} \vDash \text{ZFC} \cup \{\Psi\}.$$

Two fundamental issues remain. Assume there is a proper class of Woodin cardinals.

1. Is there a good axiom, Ψ, such that Ψ is Ω_{ZFC}-satisfiable?
2. Suppose Ψ is a sentence and that Ψ is Ω_{ZFC}-satisfiable. Ψ is *weakly good* if for all sentences ϕ in the language for the structure,

   $$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

   either the sentence

   $$(\Psi \rightarrow \text{“}\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi\text{”})$$

   is Ω_{ZFC}-valid or the sentence

   $$(\Psi \rightarrow \text{“}\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash (\neg\phi)\text{”})$$

   is Ω_{ZFC}-valid.

   Suppose that Ψ is weakly good. Must the sentence

   $$(\Psi \rightarrow \text{CH})$$

   be Ω_{ZFC}-valid?

Of course in light of the discussion above, it is evident that the Ω Conjecture settles both these questions, affirmatively.

5. Maximality Axioms

Suppose that α is an infinite ordinal such that for all ordinals $\beta < \alpha$ and for all ordinals $\gamma < \alpha$, $\beta \cdot \gamma < \alpha$. A sentence, ϕ, in the language for the structure,

$$\langle \mathcal{P}(\alpha), \alpha, \cdot, +, \in \rangle$$

is Π_2 if it is of the form, $(\forall x_0 (\exists x_1 \psi))$ where ψ is a formula with all quantifiers ranging over α (so ϕ is in essence a Π_2 sentence in the second order language for the structure $\langle \alpha, \cdot, +, \in \rangle$).

It will be necessary to consider structures,

$$\langle \mathcal{P}(\alpha), A_1, \ldots, A_n, \alpha, \cdot, +, \in \rangle,$$

where $A_1, \ldots, A_n$ are subsets of $\mathcal{P}(\alpha)$. The notion of a Π_2 sentence in the language for this structure is more awkward to define. Here it is best to pass to the structure,[5]

$$\langle H(|\alpha|^+), A_1, \ldots, A_n, \in \rangle,$$

which is logically equivalent to the structure,

$$\langle \mathcal{P}(\alpha), A_1, \ldots, A_n, \alpha, \cdot, +, \in \rangle,$$

and define ϕ to be a Π_2 sentence if it is of the form, $(\forall x_0 (\exists x_1 \psi))$ where ψ is a formula with only bounded quantifiers.

The Shoenfield Absoluteness Theorem recast in terms of Ω-logic is simply the following assertion. Suppose that ϕ is a Π_2 sentence in the appropriate language. Then either the sentence, "$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle \vDash \phi$" is Ω_{ZFC}-valid or the sentence "$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle \vDash (\neg\phi)$" is Ω_{ZFC}-valid.

This result easily generalizes to the structures,

$$\langle \mathcal{P}(\alpha), \alpha, \cdot, +, \in \rangle,$$

where $\alpha < \omega_1$.

It is straightforward to show that there exists a Π_2 sentence, ψ, such that the *Continuum Hypothesis* is false if and only if,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \psi.$$

Therefore the most one can hope to establish are results like the following: Suppose that ϕ_1 and ϕ_2 are Π_2 sentences such that both

$$\text{"}\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi_1\text{"}$$

and

$$\text{"}\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi_2\text{"}$$

are Ω_{ZFC}-consistent. Then

$$\text{"}\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash (\phi_1 \wedge \phi_2)\text{"}$$

is Ω_{ZFC}-consistent. This would show that if ϕ is a sentence which is any "simpler" than Π_2 then either the sentence, "$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi$" is Ω_{ZFC}-provable or the sentence "$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash (\neg\phi)$" is Ω_{ZFC}-provable.

The next theorem shows that not only is this true for Π_2 sentences, but a stronger version is true[6]. We require some more definitions. First we note that Ω-logic can

[5] $H(|\alpha|^+)$ denotes the transitive set consisting of all sets of hereditary cardinality less than or equal to the cardinality of α.

[6] But only for Π_2 sentences, the analogous statement for Σ_2 sentences—sentences which are expressible as the negation of a Π_2 sentence—is *false* [11].

be naturally expanded to the language with predicates for universally Baire sets (and constants for reals) [13]. So if A is universally Baire, $a \in \mathbb{R}$, and $\phi(x_0, x_1)$ is a formula in the language of set theory then one can generalize the definitions of Ω-logic to define when $\phi(a, A)$ is Ω_{ZFC}-valid and when $\phi(a, A)$ is Ω_{ZFC}-provable.

The second definition we require isolates a key new feature of the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

as compared to the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle.$$

A cofinal set $C \subseteq \omega_1$ is *closed* if for all $0 < \alpha < \omega_1$, either $\alpha \in C$ or $C \cap \alpha$ is not cofinal in α. A set $A \subset \omega_1$ is *nonstationary* if the set $\omega_1 \setminus A$ contains a subset which is a closed, cofinal, subset of ω_1. It follows from the Axiom of Choice that the collection of nonstationary subsets of ω_1 is a σ-ideal—which is an ideal closed under countable unions—denoted here as $\mathcal{I}_{\text{NS}}$. Clearly the ideal, $\mathcal{I}_{\text{NS}}$, is definable, without parameters, in the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle.$$

Theorem 9 (ZFC). *Suppose there exist a proper class of Woodin cardinals and that $A \subseteq \mathbb{R}$ is universally Baire. Suppose that ϕ_1 and ϕ_2 are Π_2 sentences such that both*

$$\text{“}\langle \mathcal{P}(\omega_1), A, \mathcal{I}_{\text{NS}}, \omega_1, \cdot, +, \in \rangle \vDash \phi_1\text{”}$$

and

$$\text{“}\langle \mathcal{P}(\omega_1), A, \mathcal{I}_{\text{NS}}, \omega_1, \cdot, +, \in \rangle \vDash \phi_2\text{”}$$

are Ω_{ZFC}-consistent. Then

$$\text{“}\langle \mathcal{P}(\omega_1), A, \mathcal{I}_{\text{NS}}, \omega_1, \cdot, +, \in \rangle \vDash (\phi_1 \wedge \phi_2)\text{”}$$

is Ω_{ZFC}-consistent. □

I now come to the main example of a sentence which is good. Let Ψ_0 be the sentence:

For each projective set A, for each Π_2 sentence ϕ, if the sentence,

$$\text{“}\langle \mathcal{P}(\omega_1), A, \mathcal{I}_{\text{NS}}, \omega_1, \cdot, +, \in \rangle \vDash \phi\text{”},$$

is Ω_{ZFC}-consistent then

$$\langle \mathcal{P}(\omega_1), A, \mathcal{I}_{\text{NS}}, \omega_1, \cdot, +, \in \rangle \vDash \phi.$$

Then Ψ_0 is good, this is the content of the next theorem.

Theorem 10 (ZFC). *Suppose there is a proper class of Woodin cardinals. Then*

1. Ψ_0 *is* Ω_{ZFC}*-consistent.*
2. *For* ***all*** *sentences* ϕ *in the language for the structure,*

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

either the sentence "$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash \phi$" *is* Ω_{ZFC}*-provable from* Ψ_0 *or the sentence*

$$\text{“}\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle \vDash (\neg\phi)\text{”}$$

is Ω_{ZFC}*-provable from* Ψ_0. □

The sentence Ψ_0 also settles the size of the continuum.

Theorem 11 (ZFC). *Suppose there is a proper class of Woodin cardinals and that* Ψ_0 *holds. Then* $c = \aleph_2$. □

In fact, the version of Ψ_0 which refers only to those projective sets $A \subseteq \mathbb{R}$ which can be defined in the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle,$$

without parameters—these are the "lightface" projective sets—is equivalent[7] to Ψ_0 and the theorem that Ψ_0 is good is true with the sentences ϕ replaced by sentences with real parameters.

The structures,

$$\langle \mathcal{P}(\alpha), \alpha, \cdot, +, \in \rangle,$$

where $\omega_1 \leq \alpha < \omega_2$ are each logically equivalent to the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle,$$

so the next structure to consider is the structure,

$$\langle \mathcal{P}(\omega_2), \omega_2, \cdot, +, \in \rangle.$$

Here maximality fails not only for Π_2 sentences, it fails for Π_1 sentences. Very likely maximality fails for Σ_1 sentences as well but this is open.

6. Conclusions

It is important to note that if there exists a proper class of Woodin cardinals then for all complete Boolean algebras, $\mathbb{B}$,

$$V^{\mathbb{B}} \vDash \Omega \text{ Conjecture}$$

if and only if the Ω Conjecture holds in V.

[7] In Ω-logic, assuming there is a proper class of Woodin cardinals [11].

Therefore it is very unlikely that the problem of the Ω Conjecture is unsolvable in the same fashion that CH is unsolvable. It could be unsolvable in the same fashion that the measure problem for the projective sets is unsolvable. But in this case the Ω Conjecture would most likely be resolved as being false.[8]

If the skeptic is going to accept the solution to the measure problem for the projective sets, but continue to claim that the problem of the *Continuum Hypothesis* is meaningless, then the skeptic must explain why the problem of the *Continuum Hypothesis* is different. The existence of good sentences (which are Ω-satisfiable) and the theorem that good sentences must imply that the *Continuum Hypothesis* is false would seem to narrow the possibilities for an argument that the problem of the *Continuum Hypothesis* is fundamentally different from the measure problem for the projective sets to essentially just the fact that neither CH or ($\neg$CH) is Ω_{ZFC}-valid. But accepting this as the reason that the problem of the *Continuum Hypothesis* is without meaning forces one to make the stronger claim that if a Π_2-sentence is not Ω_{ZFC}-valid then the claim that it is true is not meaningful (noting that both CH and ($\neg$CH) are Π_2 sentences). However the plausibility of such a claim critically depends on the Ω Conjecture.

In one extreme the Ω Conjecture is false and moreover the set

$$x = \{\phi \mid \emptyset \vDash_\Omega \phi\}$$

can be complicated as its immediate definition suggests. How might this happen? Consider the following sentence, $\Phi_0[x]$, which involves x (so in the language of set theory with a constant for x).

> x is the set $\{\phi \mid \emptyset \vDash_\Omega \phi\}$ and x is recursively equivalent to the complete Π_2 set of integers.

Assuming there is a proper class of Woodin cardinals it follows that whether or not this sentence is Ω_{ZFC}-satisfiable is generically absolute. If this sentence is Ω_{ZFC}-satisfiable then the Ω Conjecture is false (and in a very strong sense). In this case the foundational view that the only Π_2-sentences in Set Theory for which the claim of truth is meaningful are those which are Ω_{ZFC}-valid (a "many worlds" view) becomes more credible. A more persuasive case for this view could be made if for each Σ_2 sentence, ϕ, such that ϕ is Ω_{ZFC}-satisfiable, the sentence $(\phi \wedge \Phi_0[x])$ is Ω_{ZFC}-satisfiable—whether or not this is true is also generically absolute if there is a proper class of Woodin cardinals.

Of course the set,

$$x = \{\phi \mid \emptyset \vDash_\Omega \phi\},$$

could be recursively equivalent to the complete Π_2 set of integers because there is a sentence Ψ_1 such that for all Π_2 sentences ϕ,

$$V \vDash \phi$$

if and only if $\{\Psi_1\} \vDash_\Omega \phi$. While such a sentence, Ψ_1, would be compelling as a new axiom, this possibility seems rather unlikely at present. The reason is that if

[8] By a meta-mathematical argument involving hierarchies of Axioms of Infinity and *Inner Model Theory*.

such a sentence Ψ_1 exists (and holds in V) then it should follow (by introducing real parameters) that every set of reals which is ordinal definable in V, is universally Baire[9]. Finally, suppose that there exist a proper class of Woodin cardinals *and* suppose that every set of reals which is ordinal definable is universally Baire. Then the Ω Conjecture holds in HOD in the strong sense that for every sentence ϕ, if ϕ is Ω_{ZFC}-consistent then ϕ is Ω_{ZFC}-satisfiable. While this is not a formal contradiction, it seems very implausible that the Ω Conjecture could both fail in V and hold in HOD in this strong sense.

In the other extreme, the Ω Conjecture is true. But then by a corollary of the general theory of Ω-logic, the set x is definable (without parameters) in the structure,

$$\langle \mathcal{P}(c), c, \cdot, +, \in \rangle,$$

where $c = |\mathbb{R}|$. In this case the foundational view that the only Π_2-sentences in Set Theory for which the claim of truth is meaningful are those Π_2 sentences which are Ω_{ZFC}-valid is no more credible than formalism. The view rejects any genuine notion of the transfinite beyond c.

If the Ω Conjecture is false then one could simply conclude that the definition of $T \vdash_\Omega \phi$ was not correct. But if the Ω Conjecture fails badly in the sense that the sentence, $\Phi_0[x]$, involving the set x is Ω_{ZFC}-satisfiable then there is probably no reasonable notion of the length of an Ω-proof. Otherwise (assuming there is a proper class of Woodin cardinals) there would be formulas, $\psi(x_0, x_1)$ and $\phi(x_0, x_1)$, such that the following hold (regarding $x \subseteq \omega$):

1. The set

$$z = \{(i, j) \mid i \in x, j \in x, \text{ and } \psi[i, j] \text{ is } \Omega_{\mathrm{ZFC}}\text{-valid}\}$$

well-orders x;

2. For each $j \in x$,

$$\{i \mid i \in \omega \text{ and } \phi[i, j] \text{ is } \Omega_{\mathrm{ZFC}}\text{-valid}\} = \{i \mid (i, j) \notin z\}.$$

However if x is recursively equivalent to the complete Π_2 set then this cannot happen.

If the Ω Conjecture is true then in understanding why it is true, arguably our understanding of sets will have advanced considerably. We will be able to quantify the limits of forcing and we will be able to give an abstract definition of the hierarchy of *Axioms of Infinity*. Nevertheless the challenge, if the Ω Conjecture is true, will be to come up with a credible foundational view of the transfinite universe. A more specific challenge is:

Exhibit a sentence ϕ such that assertion,

$$\langle \mathcal{P}(\mathbb{R}), \mathbb{R}, +, \cdot, \in \rangle \vDash \phi,$$

is *true* but *not* $\vDash_\Omega$ valid.

[9]The "least" OD subset of $\mathbb{R}$ which is not universally Baire should be Ω-finite.

This challenge is reasonably posed *now* to those advocating a notion of truth in Set Theory which transcends mere formalism and yet maintain that the *Continuum Hypothesis* has no answer.

There are many natural candidates for ϕ other than CH and its negation, for example:

There is a Σ^2_1*-definable wellordering of* $\mathbb{R}$*;*

Every Σ^2_1*-definable set of reals is determined;*

The pointclass, Π^2_1*, has the prewellordering property;*

The pointclass, Σ^2_1*, has the prewellordering property;*

(and their negations) are each candidates (and less controversial than CH). The prewellordering property is defined in [9].

I am optimistic. Our collective experience in Set Theory to date has revealed the correct axioms for the structure,

$$\langle \mathcal{P}(\omega), \omega, \cdot, +, \in \rangle,$$

and has provided candidates for the correct axioms for the structure,

$$\langle \mathcal{P}(\omega_1), \omega_1, \cdot, +, \in \rangle.$$

I see no reason why this should stop here and I am not unduly discouraged by the fact that these structures are negligible initial segments of the universe of sets.

References

[1] Albers, Donald J., Gerald L. Alexanderson, and Constance Reid (eds.): 1990. *More Mathematical People*. Contemporary conversations. Boston, MA: Harcourt Brace Jovanovich Publishers.

[2] Dehornoy, Patrick: 2002–2003. Progrès récents sur l'hypothèse du continu [d'après Woodin]. *Séminaire Bourbaki* 55ème année: # 915.

[3] Feferman, Solomon, Harvey M. Friedman, Penelope Maddy, and John R. Steel: 2000. Does mathematics need new axioms? *Bull. Symbolic Logic* 6 (4): 401–446.

[4] Feng, Qi, Menachem Magidor, and Hugh Woodin: 1992. Universally Baire sets of reals. In: *Set Theory of the Continuum (Berkeley, CA, 1989)*, New York: Springer, 203–242.

[5] Gödel, Kurt: 1940. *The Consistency of the Continuum Hypothesis*. Princeton, N. J.: Princeton University Press.

[6] Kanamori, Akihiro: 1994. *The Higher Infinite*. Berlin: Springer.

[7] Martin, Donald A.: 1976. Hilbert's first problem: the continuum hypothesis. In: *Mathematical Developments Arising from Hilbert Problems (Proc. Sympos. Pure Math., Northern Illinois Univ., De Kalb, Ill., 1974)*, Providence, R. I.: Amer. Math. Soc., 81–92.

[8] Mitchell, William J. and John R. Steel: 1994. *Fine Structure and Iteration Trees*. Berlin: Springer.

[9] Moschovakis, Yiannis N.: 1980 . *Descriptive Set Theory*. Amsterdam: North-Holland Publishing Co..

[10] Shelah, Saharon: 1993. The future of set theory. In: *Set Theory of the Reals (Ramat Gan, 1991)*, Ramat Gan: Bar-Ilan Univ., 1–12.

[11] Woodin, W. Hugh: 1999. *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*. Berlin: Walter de Gruyter & Co., Berlin.

[12] Woodin, W. Hugh: 2001. The continuum hypothesis. I. *Notices Amer. Math. Soc.* 48 (6): 567–576.

[13] Woodin: W. Hugh: 2001. The continuum hypothesis. II. *Notices Amer. Math. Soc.* 48 (7): 681–690.

Department of Mathematics
721 Evans Hall # 3840
Berkeley, CA 94720-3840
USA

E-mail: woodin@math.berkeley.edu

A Way Out

Harvey M. Friedman

Abstract. We present a way out of Russell's paradox for sets in the form of a direct weakening of the usual inconsistent full comprehension axiom scheme, which, with no additional axioms, interprets ZFC. In fact, the resulting axiomatic theory 1) is a subsystem of ZFC + "there exists arbitrarily large subtle cardinals", and 2) is mutually interpretable with ZFC + the scheme of subtlety.

1. Newcomp

Bertrand Russell [1] showed that the Fregean scheme of full comprehension is inconsistent. Given the intuitive appeal of full comprehension (for sets), this inconsistency is known as Russell's Paradox (for sets). The modern view is to regard full comprehension (for sets) as misguided, and thereby regard Russell's Paradox (for sets) as a refutation of a misguided idea.

We first give an informal presentation of the axiom scheme investigated in this paper. Informally, the full comprehension axiom scheme in the language $L(\in)$ with only the binary relation symbol $\in$ and no equality, is, in the context of set theory,

Every virtual set forms a set.

We use the term "virtual set" to mean a recipe that is intended to be a set, but may be a "fake set" in the sense that it does not form a set. The recipes considered here are of the form $\{x : \varphi\}$, where φ is any formula in $L(\in)$.

Other authors prefer to use the term "virtual class", reflecting the idea that $\{x : \varphi\}$ always forms a class, with the understanding that x ranges over sets. Our terminology reflects the intention to consider only sets, and construct a powerful set existence axiom.

We say that $\{x : \varphi\}$ forms a set if and only if there is a set y whose elements are exactly the x such that φ. Here y must not be free in φ (and must be different from x). Thus $\{x : \varphi\}$ forms a set is expressed by

$$(\exists y)(\forall x)(x \in y \longleftrightarrow \varphi).$$

Russell showed that

$$\{x : x \notin x\} \textit{ forms a set}$$

leads to a contradiction in pure logic.

Our way out of Russell's Paradox is to modify the inconsistent Fregean scheme in this way:

Every virtual set forms a set, or _____ .

We refer to what comes after "or" as the "escape clause". The escape clause that we use involves only the extension of the virtual set and not its presentation.

We are now ready to present the comprehension axiom scheme.

Newcomp. Every virtual set forms a set, or, outside any given set, has two inequivalent elements, where all elements of the virtual set belonging to the first belong to the second.

To avoid any possible ambiguity, we make the following comments (as well as give a formal presentation in section 2).

1. For Newcomp, we use only the language $L(\in)$, which does not have equality.
2. Here "inequivalent" means "not having the same elements".
3. The escape clause asserts that for any set y, there are two unequal sets z, w in the extension of the virtual set, neither in y, such that every element of z in the extension of the virtual set is also an element of w.

We will show that

a) Newcomp is provable in ZFC + "there exists arbitrarily large subtle cardinals";

b) Newcomp is provable in ZFC + $V = L$ + SSUB, where SSUB is what we call the scheme of subtlety;

c) Newcomp and ZFC + SSUB are mutually interpretable;

d) Newcomp is interpretable in ZFC + "there exists a subtle cardinal", but Newcomp is not provable there, assuming that the latter is consistent;

e) all of the above are provable in the weak fragment of arithmetic EFA (exponential function arithmetic);

f) Newcomp and ZFC + SSUB are equiconsistent, in the sense that their consistencies are provably equivalent in EFA.

As usual, ZFC is formulated with equality; i.e., in the language $L(\in, =)$.

The interpretation of ZFC + SSUB in Newcomp presented here (or coming out of here) takes the following form (when straightforwardly adjusted). Sets in ZFC+SSUB are interpreted to be sets in Newcomp. Membership and equality between the sets in ZFC + SSUB are interpreted as two separate relations between sets in Newcomp defined by two separate formulas with exactly two free variables (no parameters). As normally required of interpretations, the usual connectives and quantifiers are

interpreted without change. Every theorem of ZFC + SSUB becomes a theorem of Newcomp when so interpreted.

There is an appropriate sense in which this interpretation is a well founded interpretation. Specifically, it is provable in Newcomp that every set has a minimal element under the above interpretation of the epsilon relation of ZFC + SSUB. We can then draw conclusions such as conservative extension results in the form: any sentence of a certain kind provable in ZFC + SSUB is provable in Newcomp. However, the statement of such results involves various coding apparatus available in Newcomp, and we do not go into this matter here. Suffice it to say that, in an appropriate sense, every arithmetical theorem of ZFC + SSUB is a theorem of Newcomp (and vice versa).

In the interpretation of Newcomp in ZFC + "there exists a subtle cardinal", the sets in Newcomp are interpreted to be some portion of the sets in ZFC + "there exists a subtle cardinal", (an initial segment, possible proper, of the constructible hierarchy), and membership between sets in Newcomp is interpreted as membership between sets in ZFC + "there exists a subtle cardinal". As normally required of interpretations, the usual connectives and quantifiers are interpreted without change. Every theorem of Newcomp becomes a theorem of ZFC + "there exists a subtle cardinal" when so interpreted.

Of course, when interpreting Newcomp in ZFC + "there are arbitrarily large subtle cardinals", we can use the identity interpretation, since Newcomp is provable there.

The system ZFC + SSUB has logical strength a shade below that of ZFC + SUB, where SUB is "there exists a subtle cardinal", and substantially stronger than the well studied large cardinal axioms weaker than SUB, such as the existence of Mahlo, weakly compact, or indescribable cardinals. In the well known Chart of Cardinals in [6]: 471, subtle fits strictly below $\kappa \rightarrow (\omega)^{<\omega}_2$, and strictly above (in logical strength) "indescribable", well within the cardinals that are compatible with $V = L$. In particular, it is provable in ZFC that

i) if κ is a subtle cardinal then there are κ cardinals $< \kappa$ which are indescribable;

ii) if $\kappa \rightarrow (\omega)^{<\omega}_2$ then there are κ cardinals $< \kappa$ that are subtle;

iii) the first subtle cardinal, if it exists, is not indescribable or even weakly compact;

iv) every subtle cardinal is n-Mahlo for all $n < \omega$.

2. Some Formalities

We let $L(\in)$ be ordinary classical first order predicate calculus with only the binary relation symbol $\in$ (no equality). We assume that x, y, z, w are distinct variables among the infinitely many variables used in $L(\in)$.

We write $z \equiv w$ for $(\forall u)(u \in z \longleftrightarrow u \in w)$.

Newcomp. Let φ be a formula of $L(\in)$ in which y, z, w do not appear. $(\exists y)(\forall x)(x \in y \longleftrightarrow \varphi) \vee (\forall y)(\exists z, w)(z, w \notin y \wedge \neg z \equiv w \wedge \varphi[x/z] \wedge \varphi[x/w] \wedge (\forall x)((\varphi \wedge x \in z) \rightarrow x \in w))$.

The following definition is used in [2] and [4]. We say that an ordinal λ is subtle if and only if

i) λ is a limit ordinal;

ii) Let $C \subseteq \lambda$ be closed and unbounded, and for each $\alpha < \lambda$ let $A_\alpha \subseteq \alpha$ be given. There exist $\alpha, \beta \in C$, $\alpha < \beta$, such that $A_\alpha = A_\beta \cap \alpha$.

It is well known that every subtle ordinal is a subtle cardinal, see [4]: 3.

We will use the following schematic form of subtlety. SSUB is the following scheme in the language of ZFC with $\in, =$. Let φ, ψ be formulas, where we view φ as carving out a class on the variable x, and ψ as carving out a binary relation on the variables x, y. Parameters are allowed in φ, ψ.

> *if φ defines a closed and unbounded class of ordinals C, and ψ defines a system $A_\alpha \subseteq \alpha$, for all ordinals α, then there exist $\alpha, \beta \in C$, $\alpha < \beta$, where $A_\alpha = A_\beta \cap \alpha$.*

This concludes the definitions that are used in the list of agenda items a)–f) at the end of section 1.

In [5] we define λ to be inclusion subtle if and only if

i) λ is a limit ordinal;

ii) Let $C \subseteq \lambda$ be closed and unbounded, and for each $\alpha < \lambda$ let $A_\alpha \subseteq \alpha$ be given. There exist $\alpha, \beta \in C, \alpha < \beta$, such that $A_\alpha \subseteq A_\beta$.

There is a corresponding scheme SISUB (scheme of inclusion subtlety).

> *if φ defines a closed and unbounded class of ordinals, C, and ψ defines a system $A_\alpha \subseteq \alpha$, for all ordinals α, then there exist $\alpha, \beta \in C$, $\alpha < \beta$, such that $A_\alpha \subseteq A_\beta$.*

Lemma 1. *An ordinal is subtle if and only if it is inclusion subtle.*

Proof. This is Theorem 1.2 of [5]. □

Theorem 2. SISUB *and* SSUB *are provably equivalent in* ZFC.

Proof. This is in clear analogy with Lemma 1. The proof is an obvious adaptation of that of Lemma 1. □

In [5] we define λ to be weakly inclusion subtle over δ if and only if

i) λ, δ are ordinals;

ii) For each $\alpha < \lambda$ let $A_\alpha \subseteq \alpha$ be given. There exists $\delta \leq \alpha < \beta$ such that $A_\alpha \subseteq A_\beta$.

There is a corresponding scheme SWISUB (scheme of weak inclusion subtlety). This is the natural principle used to prove Newcomp in section 3 (we will also use $V = L$).

> *if φ defines a system $A_\alpha \subseteq \alpha$, for all ordinals α, then there exist arbitrarily large ordinals $\alpha < \beta$ such that $A_\alpha \subseteq A_\beta$.*

Lemma 3. *The least weakly inclusion subtle ordinal over $\delta \geq 2$, if it exists, is a subtle cardinal.*

Proof. This is Theorem 1.6 of [5]. □

Theorem 4. SWISUB *and* SSUB *are provably equivalent in* ZFC + ¬SUB.

Proof. This will be an adaptation of the proof of Lemma 3. We work in SWISUB + ¬SUB. Let a closed unbounded class C of ordinals, and $A_\alpha \subseteq \alpha$, ordinals α, be appropriately given. By Theorem 2, it suffices to find $\alpha, \beta \in C$, $\alpha < \beta$, such that $A_\alpha \subseteq A_\beta$. By Lemma 3, there is no weakly inclusion subtle ordinal over 2. So for each γ, we have a counterexample $D_\alpha \subseteq \alpha$, $\alpha < \gamma$, to γ is weakly inclusion subtle over 2.

We now proceed exactly as in the proof of Theorem 1.6 in [5]. □

Corollary 5. ZFC + SSUB *and* ZFC + SWISUB *are mutually interpretable and equiconsistent* (*in the sense that the consistency statements are provably equivalent in* EFA).

Proof. The interpretation will be V if there is no subtle cardinal, and $V(\lambda)$ if there is a subtle cardinal and λ is the least subtle cardinal. □

For the interpretation of ZFC + SSUB in Newcomp, we will use another modification of SSUB that is weaker than SWISUB. We call it the scheme of very weak inclusion subtlety, written SVWISUB.

If ψ defines a system $A_\alpha \subseteq \alpha$, for all ordinals α, then there exist $2 \leq \alpha < \beta$ such that $A_\alpha \subseteq A_\beta$.

Theorem 6. SVWISUB *and* SSUB *are provably equivalent in* ZFC + ¬ SUB.

Proof. The same as for Theorem 4. In Theorem 1.6 of [5], set $\delta = 2$. □

Corollary 7. ZFC + SSUB *and* ZFC + SVWISUB *are mutually interpretable and equiconsistent* (*in the sense that the consistency statements are provably equivalent in* EFA).

Proof. Same as for Corollary 5. □

In section 5, we will interpret SVWISUB in Newcomp + Extensionality.

3. Proof and Interpretation of Newcomp

Theorem 8. Newcomp *is provable in* ZFC + $V = L$ + SWISUB. *In particular, it is provable in* ZFC + $V = L$ + SSUB, *and interpretable in* ZFC + SSUB. Newcomp *is interpretable in* ZFC + *"there exists a subtle cardinal"*.

Proof. We work in ZFC + $V = L$ + SWISUB. Let S be a proper class given by a formula with set parameters.

We construct a one-one surjective function $F : S \to \text{On}$ such that

*) for all $x, y \in S$, if $x \in y$ then $F(x) < F(y)$.

Let γ_α be the strictly increasing enumeration of all of the ranks of elements of S. Let F_0 map $S \cap V(\gamma_0 + 1)$ one-one onto an ordinal. Suppose F_β has been defined for all $0 \leq \beta < \alpha$, mapping $S \cap V(\gamma_\beta + 1)$ one-one onto an ordinal, where each function is extended by the later ones. If α is a limit ordinal, then take F_α to be the union of the F_β, $\beta < \alpha$. Suppose $\alpha = \beta + 1$. Take F_α to be an appropriate extension of F_β.

It is clear that we cannot make this construction in a definable manner; e.g., S may be all of V. However, $V = L$ provides the needed definable well ordering of V to make this construction go through.

We now define $A_\alpha \subseteq \alpha$ for all ordinals α. Recall that F is one-one onto. Take $A_\alpha = \{F(x) : x \in S \cap F^{-1}\alpha\}$, where $F^{-1}\alpha$ is the inverse of the function F at the point α. Suppose $x \in S \cap F^{-1}\alpha$. By condition *) in the construction of F, we have $F(x) < F(F^{-1}\alpha) = \alpha$. Thus we see that $A_\alpha \subseteq \alpha$.

By SWISUB, let $\alpha < \beta$ be arbitrarily large ordinals such that $A_\alpha \subseteq A_\beta$. Then $\{F(x) : x \in S \cap F^{-1}\alpha\} \subseteq \{F(x) : x \in S \cap F^{-1}\beta\}$. Since $F : S \to \text{On}$ is one-one, we have $\{x : x \in S \cap F^{-1}\alpha\} \subseteq \{x : x \in S \cap F^{-1}\beta\}$, $S \cap F^{-1}\alpha \subseteq S \cap F^{-1}\beta$. Also, since F is one-one, $F^{-1}\alpha$ and $F^{-1}\beta$ can be taken to lie outside any given set and are distinct elements of S. We have thus verified the escape clause in Newcomp.

For the second claim, obviously SWISUB is derivable from SSUB, and ZFC + $V = L$ + SSUB is interpretable in ZFC + SSUB by relativizing to L. For the third claim, obviously ZFC + SSUB is interpretable in ZFC + "there exists a subtle cardinal" by using the model $(V(\kappa), \in)$ of SSUB, where κ is the least subtle cardinal. □

Theorem 9. Newcomp *is provable in* ZFC + *"there are arbitrarily large subtle cardinals"*.

Proof. Let S be a proper class given by a formula with set parameters. We modify the proof of Theorem 8. Let the ordinals γ_α be defined as before, and let α be given. Let $\kappa > \alpha$ be a subtle cardinal.

Suppose that for some $\beta < \kappa$, the number of elements of S of rank γ_β is at least κ. Fix β to be least with this property. Note that $S \cap V(\gamma_\beta)$ has fewer than κ elements. Hence there are at least κ elements of S of rank γ_β whose intersections with S are equal. This verifies the escape clause in Newcomp with $x = V(\alpha)$.

Now suppose that for all $\beta < \kappa$, the number of elements of S of rank $\gamma_\beta < \kappa$. Then build the one-one functions f_β, $\beta < \kappa$, as in the proof of Theorem 8. This will

require only AxC, and not $V = L$. Note that $f_\kappa : S \cap V(\gamma_\kappa) \to \kappa$ is one-one onto. Proceed to verify the escape clause in Newcomp with $x = V(a)$, using the subtlety of κ, as in the proof of Theorem 8. □

Theorem 10. Newcomp *is not provable in* ZFC + $V = L$ + *"there exists a subtle cardinal", assuming the latter is consistent.* Newcomp *is not provable in* ZFC *together with any existential sentence in the language of set theory with equality and the power set operation, with bounded quantifiers allowed, assuming the latter is consistent.*

Proof. Let M be a model of ZFC + $V = L$ + "there exists a subtle cardinal". Let κ be any subtle cardinal in the sense of M. Let M' be the same as M if M satisfies "there is no strongly inaccessible cardinal $> \kappa$"; otherwise M' is the restriction of M to the sets of rank less than the first inaccessible cardinal $> \kappa$ in the sense of M. Then M' satisfies ZFC + $V = L$.

In M', we construct a definable assignment $A_\alpha \subseteq \alpha$, ordinals α, as follows. If $\alpha > \kappa$ is a successor ordinal, let $A_\alpha = \{0, \alpha - 1\}$. If $\alpha > \kappa$ is a limit ordinal of cofinality $< \alpha$, let A_α be an unbounded subset of α of order type $cf(\alpha)$ whose first two elements are $1, cf(\alpha)$. If $\alpha > \kappa$ is the next cardinal after the cardinal β, let $A_\alpha = \{2\} \cup [\beta, \alpha)$. For $\alpha \leq \kappa$, let $A_\alpha = \alpha$. Note that the strict sup of every A_α is α.

We claim that $\kappa < \alpha < \beta$ and $A_\alpha \subseteq A_\beta$ is impossible. If this holds then either α,β are both successor ordinals, or α,β are both nonregular limit ordinals, or α,β are both successor cardinals. The first and third cases are dispensed with immediately. For the second case, we have $cf(\alpha) = cf(\beta)$ and A_α, A_β have order type $cf(\alpha) < \alpha$. Since A_α has strict sup α and A_β has strict sup $\beta > \alpha$, this is impossible.

We can now convert this construction to a counterexample to Newcomp. The conversion is according to the proof of Theorem 2.5, iii) → i), in [5], with $\lambda = \text{On}$ and $\delta = \kappa$. For the convenience of the reader, we repeat the relevant part from there with $\lambda = \text{On}$ and $\delta = \kappa$, in the next three paragraphs, working within M'.

We define $f : \text{On} \to V$ as follows. $f(\alpha) = \{f(\beta) : \beta \in A_\alpha\}$. By transfinite induction, each $f(\alpha)$ has rank α. Let S be the range of f. Then S is a transitive set of rank λ, where S has exactly one element of each rank $< \lambda$.

We claim that $f(\alpha) \subseteq f(\beta) \to A_\alpha \subseteq A_\beta$. To see this, suppose $f(\alpha) \subseteq f(\beta)$. Then $\{f(\mu) : \mu \in A_\alpha\} \subseteq \{f(\mu) : \mu \in A_\beta\}$. Since f is one-one, $A_\alpha \subseteq A_\beta$.

Suppose there exists a 2 element chain $\{x \subsetneq y\} \subseteq S$, $\text{rk}(x), \text{rk}(y) \geq \kappa$. Since $x \subsetneq y$, we have $\text{rk}(x) \leq \text{rk}(y)$, and hence $\kappa \leq \text{rk}(x) < \text{rk}(y)$. Write $\text{rk}(x) = \alpha$ and $\text{rk}(y) = \beta$. Then $x = f(\alpha)$ and $y = f(\beta)$. Therefore $A_\alpha \subseteq A_\beta$. Since $\kappa \leq \alpha < \beta$, this contradicts the choice of (A_α).

We now claim that S provides a counterexample to Newcomp in M'. Clearly S is a proper class in M'. However, there are no $x, y \in S$, $x \neq y$, outside of $V(\kappa)$, with $S \cap x \subseteq y$.

For the second claim, we again begin with a model M of ZFC + $(\exists x_1, \ldots, x_k)\ (\varphi)$ where φ is a bounded formula in $\in$, $=$, and the power set operation. Fix a limit cardinal κ in M such that $(\exists x_1, \ldots, x_k \in V(\kappa))(\varphi)$ holds in M. Choose a well ordering X of $V(\kappa)$ in M, and pass to the submodel $L[X]$ of M. Finally, let M' be $L[X]$ if there is

no strongly inaccessible cardinal above κ in the sense of $L[X]$; otherwise M' is the result of chopping off at the first strongly inaccessible cardinal above κ in the sense of $L[X]$. Then M' satisfies ZFC + $(\exists x_1, \ldots, x_k)(\varphi)$. Now repeat the above argument for the first claim to show that M' does not satisfy Newcomp. □

4. Newcomp + Ext in Newcomp

In this section, we give an interpretation of Newcomp + Extensionality in Newcomp.

Let us be precise about what these two theories are. Recall that Newcomp is formulated as a scheme in $L(\in)$; i.e., without equality. See section 2.

However, we formulate Newcomp + Ext in $L(\in, =)$; i.e., with equality. This is convenient for section 5. Let x, y, z, w be distinct variables throughout this section. The axioms of Newcomp+Ext, in addition to the logical axioms for $L(\in, =)$, including the equality axioms for $L(\in, =)$, are

Extensionality. $(\forall z)(z \in x \longleftrightarrow z \in y) \rightarrow x = y$.

Newcomp. Let φ be a formula of $L(\in, =)$ in which y, z, w do not appear. $(\exists y)(\forall x)(x \in y \longleftrightarrow \varphi) \vee (\forall y)(\exists z, w)(z, w \notin y \wedge \neg z = w \wedge \varphi[x/z] \wedge \varphi[x/w] \wedge (\forall x)((\varphi \wedge x \in z) \rightarrow x \in w))$.

We first give the interpretation (and its verification), where a few points remain to be handled formally, as noted. We follow this by a formal treatment of the few remaining points.

In Newcomp, we define $x \equiv y$ if and only if $(\forall z)(z \in x \longleftrightarrow z \in y)$.

Sitting in Newcomp, we call a set x *extensional* if and only if for any finite sequence $x_1 \in x_2 \in \cdots \in x_k = x$, $k \geq 2$, for all y, we have $x_1 \equiv y \rightarrow y \in x_2$. This is a place where we need to fill in some details, because we don't have natural numbers and finite sequences readily available in Newcomp.

We interpret the sets for Newcomp + Ext to be the extensional sets. Membership is interpreted as membership. Equality is interpreted as $\equiv$ in Newcomp.

We first check that the interpretations of the equality axioms of Newcomp + Ext are provable in Newcomp.

For the equality axioms, it remains to prove the interpretations of

$x = y \rightarrow (z \in x \rightarrow z \in y)$

$x = y \rightarrow (x \in z \rightarrow y \in z)$.

Let x, y, z be extensional. Assume the interpretation of $x = y$. This is $(\forall w)(w \in x \longleftrightarrow w \in y)$. Obviously $z \in x \rightarrow z \in y$. We need to check that $x \in z \rightarrow y \in z$. Assume $x \in z$. We apply the definition of extensional to $x \in z$. Since $x \equiv y$, we have $y \in z$.

We now check that the interpretation of Ext is provable in Newcomp.

Extensionality reads

$(\forall z)(z \in x \longleftrightarrow z \in y) \rightarrow x = y$.

The first crucial Lemma we need is that every element of an extensional set is extensional. This is informally clear, but we will give a careful proof in Newcomp later.

Let x, y, z be extensional. Assume the interpretation of $(\forall z)(z \in x \longleftrightarrow z \in y)$. Then for all extensional z, $z \in x \longleftrightarrow z \in y$. Now let z be arbitrary. If $z \in x$ then z is extensional, and hence $z \in y$. If $z \in y$ then z is extensional, and hence $z \in x$. Thus we conclude that $x \equiv y$, which is the interpretation of $x = y$.

We now come to the Newcomp axiom scheme. We want to interpret the universal closure of the Newcomp axiom.

1) $\{x : \varphi\}$ forms a set or, outside any given set, there exist y, z, $\neg y = z$, $\varphi[x/y]$, $\varphi[x/z]$, with $(\forall x \in y)(\varphi \rightarrow x \in z)$

where y, z do not appear in φ. Let $v_1, \ldots, v_k$ be a complete list of parameters in this Newcomp axiom, without repetition.

The interpretation of the universal closure of the above Newcomp axiom is the sentence

2) For all extensional $v_1, \ldots, v_k$, (∃ extensional y)(∀ extensional x)($x \in y \longleftrightarrow \varphi^*$) or, for any extensional w, there exist extensional $y, z \notin w$, with $\neg y \equiv z$, $\varphi^*[x/y]$, $\varphi^*[x/z]$, and (∀ extensional $x \in y$)($\varphi^* \rightarrow x \in z$)

where φ^* is the result of relativizing all quantifiers in φ to the extensional sets, and replacing $=$ with $\equiv$. Here y, z, w do not appear in φ and $v_1, \ldots, v_k, x, y, z, w$ are distinct.

We can rewrite 2) in Newcomp as

3) For all extensional $v_1, \ldots, v_k$, (∃ extensional y)(∀ extensional x)($x \in y \longleftrightarrow \varphi^*$) or, for any extensional w, there exist extensional $y, z \notin w$, with $\neg y \equiv z$, $\varphi^*[x/y]$, $\varphi^*[x/z]$, and $(\forall x \in y)(\varphi^* \rightarrow x \in z)$.

We will actually prove the following strengthening of 3) in Newcomp. Let φ' be $\varphi^* \wedge x\ is$ extensional.

4) For all extensional $v_1, \ldots, v_k$, (∃ extensional y)$(\forall x)(x \in y \longleftrightarrow \varphi')$ or, for any w, there exists $y, z \notin w$, with $\neg y \equiv z$, $\varphi'[x/y]$, $\varphi'[x/z]$, and $(\forall x \in y)(\varphi' \rightarrow x \in z)$.

Let us verify that 4) $\rightarrow$ 3) in Newcomp. Clearly $(\forall x)(x \in y \longleftrightarrow \varphi')$ implies (∀ extensional x)($x \in y \longleftrightarrow \varphi^*$) since if x is extensional, then $\varphi' \longleftrightarrow \varphi^*$. Also, any such y, z in 4) are extensional because of the construction of φ'. In addition, $(\forall x \in y)(\varphi' \rightarrow x \in z)$ implies $(\forall x \in y)(\varphi^* \rightarrow x \in z)$. To see this, assume $(\forall x \in y)(\varphi' \rightarrow x \in z)$, and let $x \in y$ and φ^*. Since y is extensional, x is extensional, and so φ', and hence $x \in z$.

Note that we have just used the following first crucial lemma: every element of an extensional set is extensional.

Also note that 4) is almost in the form of the universal closure of a Newcomp axiom. The problem is the displayed existential quantifier. In Newcomp we obviously have

5) For all extensional $v_1, \ldots, v_k$, $(\exists y)(\forall x)(x \in y \longleftrightarrow \varphi')$ or, for any w, there exists $y, z \notin w$, with $\neg y \equiv z$, $\varphi'[x/y]$, $\varphi'[x, z]$, and $(\forall x \in y)(\varphi' \rightarrow x \in z)$.

It remains to prove in Newcomp,

6) For all extensional $v_1,\dots,v_k$, $(\exists y)(\forall x)(x \in y \longleftrightarrow \varphi') \to (\exists$ extensional $y)$ $(\forall x)(x \in y \longleftrightarrow \varphi')$.

We first claim that

7) For all extensional $v_1, \dots, v_k$, $(\forall x, u)(x \equiv u \to (\varphi' \longleftrightarrow \varphi'[x/u]))$

where u is not free in φ' and not among $v_1, \dots, v_k$.

To prove 7), first note that we have

8) For all extensional $v_1, \dots, v_k$, $(\forall x, u)(x = u \to (\varphi \longleftrightarrow \varphi[x/u]))^*$

since this is the interpretation of a theorem of logic, and we have secured the interpretation of the equality axioms. Hence we have

9) For all extensional $v_1, \dots, v_k$, $(\forall x, u)(x \equiv u \to (\varphi^* \longleftrightarrow \varphi^*[x/u]))$.

For the next paragraph, we will need a second crucial Lemma. If x is extensional and $x \equiv y$, then y is extensional. This is informally clear, but we will give a careful proof in Newcomp later.

To finish the proof of 7), let $v_1, \dots, v_k$ be extensional and let $x \equiv u$. If φ' then x is extensional, and so u is extensional. Also φ^*, and so by 9), $\varphi^*[x/u]$. Hence $\varphi'[x/u]$. For the other direction, if $\varphi'[x/u]$ then u is extensional, and so x is extensional. Also $\varphi^*[x/u]$, and so φ^*. Hence φ'. This verifies 7).

We now complete the verification of 6) in Newcomp. Let $v_1, \dots, v_k$ be extensional, and let $(\forall x)(x \in y \longleftrightarrow \varphi')$. Then y is a set of extensional sets. By 9), y is a set of extensional sets, where any set equivalent to an element of y is an element of y. It is informally clear that y itself is extensional.

Thus we need a third crucial Lemma. If x is a set of extensional sets, and $(\forall y, z)((y \in x \wedge y \equiv z) \to z \in x)$, then x is extensional.

This completes the verification of the interpretation of Newcomp+Ext in Newcomp.

We now formally treat extensional sets and the three crucial Lemmas used above. We will work entirely within Newcomp.

Lemma 11. *Separation holds. I.e.,* $(\exists y)(\forall x)(x \in y \longleftrightarrow (x \in a \wedge \varphi))$*, where* y *is not free in* φ.

Proof. $\{x : x \in a \wedge \varphi\}$ does not have two inequivalent elements outside a. So it forms a set. □

By Lemma 11, we can use the notation $y \equiv \{x : x \in a \wedge \varphi\}$ for $(\forall x)(x \in y \longleftrightarrow (x \in a \wedge \varphi))$. This determines x up to equivalence.

Lemma 12. *For any* x, y*, there exists* z *such that* $(\forall w)(w \in z \longleftrightarrow (w \in x \vee w \equiv y))$.

Proof. $\{w : w \in x \vee w \equiv y\}$ does not have two inequivalent elements outside x. So it forms a set. □

By Lemma 12, we can use the notation $z \equiv x \cup \{y\}$ for $(\forall w)(w \in z \longleftrightarrow (w \in x \vee w \equiv y))$. This determines z up to equivalence. We also use the notation $z \equiv \{y\}$ for $(\forall w)(w \in z \longleftrightarrow w \equiv y)$, which also determines z up to equivalence.

Recall that, informally in Newcomp, a set x is extensional if and only if for any finite sequence $x_1 \in x_2 \in \cdots \in x_k = x$, $k \geq 2$, for all y, if $x_1 \equiv y$ then $y \in x_2$.

To formalize this, we use the notion of an x-set, for any set x.

An epsilon closed subset of a set b is $w \subseteq b$ such that any element of an element of w that lies in b lies in w.

An *x-set* is a set b such that

i) All sets equivalent to x lie in b;

ii) Every epsilon closed subset of b containing all sets equivalent to x, is equivalent to b.

We say that x is 1-*extensional* if and only if every set equivalent to an element of x is an element of x.

We say that x is *extensional* if and only if every element of every x-set is 1-extensional.

Note that it is provable in ZF that the x-sets are exactly the sets b such that all sets equivalent to x lie in b, and every element of b is the first term of a finite epsilon chain of length ≥ 1 ending with a set equivalent to x. Thus our formal definition of extensional matches the informal definition of extensional.

Lemma 13. *Let b be an x-set. Assume $y \equiv \{x\}$. Then any $c \equiv b \cup \{y\}$ is a y-set.*

Proof. Clearly every set equivalent to y lies in c. Let z be an epsilon closed subset of c containing all sets equivalent to y. Then all sets equivalent to x lie in z. Let $w \equiv z \cap b$. We claim that w is an epsilon closed subset of b containing all sets equivalent to x. To see this, let $v \in u \in w$, $v \in b$. Then $v \in u \in z$ and $v \in c$, and so $v \in z$, and hence $v \in w$. This establishes the claim.

Since b is an x-set, $w \equiv b$. It remains to show that z contains all elements of c that are not in b. I.e., z contains all sets equivalent to y. This we already have. □

Lemma 14. *Every element of an extensional set is extensional.*

Proof. Let $x \in y$, where y is extensional. Every element of every y-set is 1-extensional. Let u be an element of some x-set b. Let $y \equiv \{x\}$. Let $c \equiv b \cup \{y\}$. By Lemma 13, c is a y-set and $u \in c$. Hence u is 1-extensional. □

Lemma 15. *If $x \equiv y$ then every x-set is a y-set.*

Proof. Let b be an x-set. We verify that b is a y-set. Clearly every set equivalent to y lies in b. Suppose z is an epsilon closed subset of b containing all sets equivalent to y as an element. Then z is an epsilon closed subset of b containing all sets equivalent to x as an element, and so $z \equiv b$. □

Lemma 16. *If x is extensional and $x \equiv y$, then y is extensional.*

Proof. Suppose z is an element of some y-set. By Lemma 15, z is an element of some x-set, and so z is 1-extensional. □

Lemma 17. *Let b be an x-set and y be an element of an element of b. Then there is an x-set c such that $y \in c$.*

Proof. Let b, x, y be as given. Let $y \in z \in b$. Let $d \equiv b \cup \{y\}$, and $c \equiv \{z \in d : z \in b \vee z$ is an element of an element of $b\}$. Clearly $b \subseteq c$ and $y \in c$.

Suppose w is an epsilon closed subset of c containing all sets equivalent to x. Let $w' \equiv w \cap b$. We claim that w' is an epsilon closed subset of b containing all sets equivalent to x. To see this, let $v \in u \in w'$, $v \in b$. Then $u, v \in c$, and so $v \in w'$. This establishes the claim. Hence $w' \equiv b$.

It remains to verify that $c \subseteq w$. The elements of c that are not in b are elements of elements of b. Since w is an epsilon closed subset of c, the elements of c not in b lie in w. □

Lemma 18. *Let $b \equiv \{u\}$. Then b is a u-set.*

Proof. Clearly b contains all sets equivalent to u. Let z be an epsilon closed subset of b containing all sets equivalent to u. Then obviously $z \equiv b$. □

Lemma 19. *Let b be an x-set, $y \in b$, $\neg y \equiv x$. There exists $z \in x$ such that y is an element of some z-set.*

Proof. Let w be the set of all elements of b that lie in some z-set, $z \in x$, together with the sets equivalent to x. It suffices to show that $b \subseteq w$. For this it suffices to show that w is an epsilon closed subset of b containing all sets equivalent to x. I.e., w is an epsilon closed subset of b.

Let $u \in v \in w$, $u \in b$. First suppose $v \equiv x$. Let $c \equiv \{u\}$. By Lemma 18, c is a u-set, $u \in x$, with $u \in c$. Hence $u \in w$.

Now suppose $\neg v \equiv x$. Then v lies in some z-set, $z \in x$. By Lemma 17, since $u \in v$, we see that u lies in some z-set, $z \in x$. Hence $u \in w$. □

Lemma 20. *Let x be a* 1*-extensional set of extensional sets. Then x is extensional.*

Proof. Let x be as given, and let y be an element of some x-set b. If $y \equiv x$ then y is 1-extensional. If $\neg y \equiv x$ then by Lemma 19, y is an element of some z-set with $z \in x$. So y is 1-extensional. □

Theorem 21. Newcomp + Extensionality *is interpretable in* Newcomp.

Proof. By the argument given before the Lemmas. The three crucial Lemmas that were cited are Lemmas 14, 16, and 20, respectively. □

5. ZFC + V = L + SVWISUB in Newcomp

In light of Theorem 21, we have only to interpret ZFC + $V = L$ + SVWISUB in Newcomp + Extensionality.

We will use the term "virtual set" whenever we have a definable property of sets presented as $\{x : \varphi\}$, with parameters allowed.

Until the proof of Lemma 60 is complete, we will work entirely within the system Newcomp + Ext.

We will use abstraction notation with braces to indicate virtual sets. We say that these expressions exist to indicate that they form sets.

All lower case letters will represent sets (not virtual sets).

It is convenient to adopt the following terminology. Let A be a virtual set. An expansion in A consists of two elements $x, y \in A$ such that $x \neq y$ and $x \cap A \subseteq y$. We say that an expansion lies outside z if and only if neither component is a member of z.

We say that an expansion meets z if and only if at least one component is a member of z.

Obviously, we can reformulate Newcomp in these two ways:

If all expansions in a virtual set meet some given set, then the virtual set forms a set.

If no expansion in a virtual set lies outside some given set, then the virtual set forms a set.

Lemma 22. *Any virtual subset of a set forms a set. The empty set exists. The intersection of any virtual set with a set forms a set. For all x, y, $x \cup \{y\}$ exists. For all $k \geq 0$ and $x_1, \ldots, x_k$, $\{x_1, \ldots, x_k\}$ exists.*

Proof. The first claim is obvious since all expansions of the virtual subset lie in the set. The second and third claims follow immediately from the first claim. For the fourth claim, note that every expansion in $x \cup \{y\}$ meets x. The fifth claim follows from the second and fourth claims. □

Lemma 23. *Let A be a transitive virtual set. The expansions in A are exactly the $x, y \in A$ with $x \subsetneq y$.*

Proof. For the forward direction, let $x, y \in A, x \neq y, x \cap A \subseteq y$. Then $x \subseteq y$ by the transitivity of A. The reverse direction is immediate. □

We write $\langle y, z \rangle = \{\{y\}, \{y, z\}\}$.

Lemma 24. *$x \cup \{\{y\} : y \in x\}$ exists. $x \cup \{\{y, z\} : y, z \in x\}$ exists. $x \cup \{\{y, z\} : y, z \in x\} \cup \{\{\{y\}\} : y \in x\}$ exists. $x \cup \{\{y, z\} : y, z \in x\} \cup \{\langle y, z \rangle : y, z \in x\}$ exists.*

Proof. Observe that all four virtual sets are transitive, and so we can apply Lemma 23. For the first claim, note that no expansions lie outside x. For the second claim, note that no expansions lie outside $x \cup \{\{y\} : y \in x\}$. For the third claim, note that no expansions lie outside $x \cup \{\{y, z\} : y, z \in x\}$. For the fourth claim, we show that no expansions lie outside $x \cup \{\{y, z\} : y, z \in x\} \cup \{\{\{y\}\} : y \in x\}$. Let $\langle a, b\rangle$, $\langle c, d\rangle$ be an expansion in $x \cup \{\{y, z\} : y, z \in x\} \cup \{\langle y, z\rangle : y, z \in x\}$ outside $x \cup \{\{y, z\} : y, z \in x\} \cup \{\{\{y\}\} : y \in x\}$. Then $\{\{a\}, \{a, b\}\} \subseteq \{\{c\}, \{c, d\}\}$, $a \neq b$, $c \neq d$. Hence the left and right sides have two elements, and each consist of a set with one element and a set with two elements. So $\{a\} = \{c\}$ and $\{a, b\} = \{c, d\}$. Hence $a = c$ and $b = d$. □

We write $x \cdot x$ for $\{\langle y, z\rangle : y, z \in x\}$. A binary relation on x is a subset of $x \cdot x$. We often use R for binary relations, and write $R(a, b)$ for $\langle a, b\rangle \in R$.

Note that by Lemma 22, all virtual binary relations on x exist (as sets). Using $(x \cdot x) \cdot (x \cdot x)$, we can simulate all virtual 4-ary relations on x as sets in the obvious way. We can continue doubling the arity in this way, thus making available all Cartesian powers of x by obvious simulation. This is very powerful, in light of separation (Lemma 22), and means that, in an appropriate sense, we have full second order logic over $(x, \in)$ at our disposal.

We caution the reader that the existence of $x \cdot y$ is apparently not available in Newcomp + Ext. Even the more fundamental $x \cup y$ is apparently not available in Newcomp + Ext.

We say that x is well founded if and only if for all nonempty $y \subseteq x$, there exists $z \in y$ such that z has no elements in common with y.

We define $W(x)$ if and only if x is transitive and well founded.

We wish to extract an "ordinal" out of the x with $W(x)$. Of course, we can't construct anything like von Neumann ordinals. Instead, we develop the usual rank comparison relation on x, and use its equivalence classes to simulate ordinals.

We define $C(R, x, y)$ if and only if

i) $W(x)$, $W(y)$, and $x \cdot y$ exists;

ii) $R \subseteq x \cdot y$;

iii) for all $z \in x$ and $w \in y$, $R(z, w)$ if and only if $(\forall a \in z)(\exists b \in w)(R(a, b))$.

Lemma 25. *There is at most one $R \subseteq x \cdot y$ with $C(R, x, y)$.*

Proof. Let $C(R, x, y)$, $C(R', x, y)$, $R \neq R'$. Then $W(x)$, $W(y)$, and $x \cdot y$ exists. Let $z \in x$ be epsilon least such that $(\exists w \in y)(R(z, w) \longleftrightarrow \neg R'(z, w))$. Then $(\forall a \in z)(\forall b \in y)(R(a, b) \longleftrightarrow R'(a, b))$. Hence $(\forall a \in z)(\exists b \in w)(R(a, b)) \longleftrightarrow (\forall a \in z)(\exists b \in w)(R'(a, b))$. Hence $R(z, w) \longleftrightarrow R'(z, w)$. □

Lemma 26. *Let $C(R, x, y)$ and $z \subseteq y$ be transitive. Then $C(R \cap x \cdot z, x, z)$.*

Proof. By Lemma 22, $W(z)$ and $x \cdot z$ exists. Let $u \in x$ and $v \in z$. Then $R(u, v)$ if and only if $(\forall a \in u)(\exists b \in v)(R(a, b) \wedge \langle a, b\rangle \in x \cdot z)$. □

Lemma 27. *Let $C(R, x, y)$, $C(R', x, z)$. Then $(\forall z \in x)(\forall w \in y \cap z)(R(z, w) \longleftrightarrow R'(z, w))$.*

Proof. By Lemma 26, $C(R \cap x \cdot z, x, z)$, $C(R' \cap x \cdot z, x, z)$. By Lemma 25, $R \cap x \cdot z = R' \cap x \cdot z$. □

Lemma 28. *Let $W(x)$, $W(y)$, and $x \cdot y$ exist. Let A be a virtual set of transitive subsets of y such that for all $z \in A$, there exists R with $C(R, x, z)$. Then there exists R' with $C(R', x, \cup A)$.*

Proof. First of all, $\cup A$ is a transitive subset of y, so $W(\cup A)$. By Lemma 25, for all $z \in A$, there exists a unique R_z with $C(R_z, x, z)$. Since these R_z are subsets of $x \cdot y$, we can set R' to be the union of the R_z, and know that R' exists. Let $a \in x$ and $b \in \cup A$. We verify that $R'(a, b) \longleftrightarrow (\forall c \in a)(\exists d \in b)(R'(a, b))$. Let $b \in z \in A$. Then $R_z(a, b) \longleftrightarrow (\forall c \in a)(\exists d \in b)(R_z(a, b))$. By Lemma 27, for all $a \in x$ and $b \in z$, $R_z(a, b) \longleftrightarrow R'(a, b)$. □

Lemma 29. *Let $W(x)$, $W(y)$, and $x \cdot y$ exist. There exists a unique R such that $C(R, x, y)$. In particular, if $W(x)$ then there exists a unique R such that $C(R, x, x)$.*

Proof. Uniqueness is by Lemma 25. For existence, let x, y be as given. We first claim that every $z \in y$ is an element of a transitive set $w \subseteq y$ such that for some R, $C(R, x, w)$. Suppose this is false, and let $z \in y$ be an epsilon minimal counterexample.

Now every $w \in z$ is an element of a transitive set $u \subseteq y$ such that for some R, $C(R, x, u)$. Let z^* be the union of all transitive $u \subseteq y$ such that for some R, $C(R, x, u)$. By Lemma 28, $z \subseteq z^*$ and z^* is a transitive subset of y and for some R, $C(R, x, z^*)$. Fix such an R. Now $z^* \cup \{z\}$ is a transitive subset of y. Define $R'(a, b) \longleftrightarrow (R(a, b) \vee (b = z \wedge (\forall c \in a)(\exists d \in b)(R(c, d))))$. Then $C(R', x, z^* \cup \{z\})$. This contradicts the choice of z.

We have thus shown that every $z \in y$ is an element of a transitive set $w \subseteq y$ such that for some R, $C(R, x, w)$. To complete the proof, apply Lemma 28.

The second claim is from the first claim and Lemma 24. □

Lemma 30. *Let $C(R, x, x)$. Then R is reflexive, transitive, connected.*

Proof. Let $C(R, x, x)$. Suppose R is not reflexive, and let $y \in x$ be epsilon minimal such that $\neg R(y, y)$. Then $R(x, x)$ is immediate.

Suppose R is not transitive, and let y, z, w be a counterexample to transitivity, with y chosen to be epsilon minimal. Assume $(\forall a \in y)(\exists b \in z)(R(a, b))$, $(\forall a \in z)(\exists b \in w)(R(a, b))$. Let $a \in y$. Fix $b \in z$, $R(a, b)$. Fix $c \in w$, $R(b, c)$. By the minimality of y, $R(a, c)$.

Suppose R is not connected, and let y, z be a counterexample to connectivity with y chosen to be epsilon minimal. Now $\neg(\forall a \in y)(\exists b \in z)(R(a,b))$, $\neg(\forall a \in z)(\exists b \in y)(R(a,b))$. Let $a \in y$, $(\forall b \in z)(\neg R(a,b))$. Let $b \in z$, $(\forall c \in y)(\neg R(b,c))$. Then $\neg R(a,b)$, $\neg R(b,a)$. This contradicts the minimality of y. □

Let $W(x)$. We write $\leq(x)$ for the unique R such that $C(R,x,x)$. We write $\equiv(x)$ for the relation $\leq(x)(a,b) \wedge \leq(x)(b,a)$. We write $<(x)$ for the relation $\leq(x)(a,b) \wedge \neg \leq(x)(b,a)$. Analogously for $>(x)$, $\geq(x)$.

The mere appearance of any of the four expressions defined in the previous paragraph will be taken to imply that $W(x)$.

Lemma 31. *If $W(x)$, $a \in b \in x$, then $<(x)(a,b)$. If $W(x)$ and x is nonempty then x has the epsilon minimum element $\emptyset$ and $\leq(x)$ has the unique minimum element $\emptyset$. If $W(x)$ and x has more than one element then $\leq(x)$ without $\emptyset$ has the unique minimum element $\{\emptyset\}$.*

Proof. Let a, b be a counterexample where a is epsilon minimal. We first show that $\leq(x)(a,b)$. Let $c \in a$. Then $<(x)(c,a)$. Hence $(\forall c \in a)(\exists d \in b)(\leq(x)(c,d))$. Therefore $\leq(x)(a,b)$.

Now suppose $\leq(x)(b,a)$. Then $(\forall c \in b)(\exists d \in a)(\leq(x)(c,d))$. Let $d \in a$, $\leq(x)(a,d)$. This contradicts the epsilon minimality of a.

Now let $W(x)$ and x be nonempty. Let u be an epsilon minimal element of x. Since x is transitive, $u = \emptyset$. Obviously, $\emptyset$ is $\leq(x)$ minimum. Also if $v \in x$ is nonempty, then $<(x)(\emptyset, x)$.

Now assume that x has more than one element. Let v be an epsilon minimal element of $x \backslash \{\emptyset\}$. Since x is transitive, the only element of v is $\emptyset$, and so $v = \{\emptyset\}$. Also $\leq(\{\emptyset\}, u)$ for all $u \in x \backslash \{\emptyset\}$, since $\leq(\emptyset, w)$ for all $w \in x$. Let $b \in x$, $b \neq \emptyset, \{\emptyset\}$. Let $c \in b$, $c \neq \emptyset$. Then $\leq(x)(b, \{\emptyset\})$ is impossible since $\neg \leq(x)(c, \emptyset)$. Hence the uniqueness of $\{\emptyset\}$ as a minimum element of $\leq(x)$ without $\emptyset$ is established. □

Lemma 32. *Assume $W(x)$, $a, b \in x$. $<(x)(a,b)$ if and only if $(\exists d \in b)(\leq(x)(a,d))$.*

Proof. Let $a, b \in x$ be a counterexample to the forward direction, where a is epsilon minimal. We have $<(x)(a,b)$, $\neg(\exists d \in b)(\leq(x)(a,d))$.

Now $\neg \leq(x)(b,a)$. Hence $\neg(\forall d \in b)(\exists c \in a)(\leq(x)(b,a))$. Fix $d \in b$ such that $(\forall c \in a)(<(x)(c,d))$. We need only show that $\leq(x)(a,d)$. Let $c \in a$. Then $<(x)(c,d)$. By the epsilon minimality of a, let $\leq(x)(c,e)$, $e \in d$. We have thus shown that $(\forall c \in a)(\exists e \in d)(\leq(x)(c,e))$. This verifies that $\leq(x)(a,d)$.

For the reverse direction, let $d \in b$, $\leq(x)(a,d)$. By Lemma 31, $<(x)(d,b)$. Hence $<(x)(a,b)$. □

Lemma 33. *If $W(x)$ then $<(x)$ is well founded. $\leq(x)$ is connected.*

Proof. Let u be a nonempty subset of x. Let $y \in x$ be epsilon minimal such that $(\exists b \in u)(\geq(x)(y, b))$. Let $<(x)(z, y)$. We claim that $z \notin u$. Suppose $z \in u$. By Lemma 32, let $\leq(x)(z, w)$, $w \in y$. This contradicts the epsilon minimality of y.

We have shown that no $z \in u$ has $<(x)(z, y)$. We would be done if $y \in u$. However, let $b \in u$, $\geq(x)(y, b)$. We can rule out $>(x)(y, b)$ since by Lemma 32, some element of y would dominate b under $\geq(x)$. Hence $\equiv(x)(y, b)$. Therefore $b \in u$ is $<(x)$ minimal as required.

Since $C(\leq(x), x, x)$, by Lemma 30, $\leq(x)$ is connected. □

This completes our treatment of "ordinals" in Newcomp + Ext. Specifically, we use the $\leq(x)$, $<(x)$, $\equiv(x)$, for x with $W(x)$.

A major difficulty arises because we apparently cannot prove the existence of $x \cup y$, even if $W(x)$ and $W(y)$. Because of this, we don't have flexibility in constructing relations between different sets.

We have discovered a way of comparing $\leq(x)$ and $\leq(y)$ for at least the relevant x, y, even though we don't have $x \cup y$. This will be good enough for our purposes.

An *x-system* is an $S \subseteq x \cdot x$ such that

i) $W(x)$;

ii) $(S(a, b) \wedge \equiv(x)(a, c) \wedge \equiv(x)(b, d)) \rightarrow S(c, d)$;

iii) $S(a, b) \rightarrow <(x)(b, a)$;

iv) the least strict upper bound of the b's such that $S(a, b)$ is given by a;

v) for all a, b, if $<(x)(a, b)$ and $(\forall y)(S(a, y) \rightarrow S(b, y))$, then $a = \emptyset$ or $a = \{\emptyset\}$.

In order to get a clear understanding of x-systems, look at the cross sections $\{b : S(a, b)\}$. These cross sections are essentially sets of "ordinals" with "strict sup" a. They have the property that the only proper inclusions between them are the cross sections of $\emptyset$, $\{\emptyset\}$, which are the "0" and "1" of $<(x)$. In other words, from the point of view of full set theory, an x-system is a counterexample to "rk(x) is weakly inclusion subtle over 2", in the sense of [5], using the elements of x under the equivalence relation "having the same rank", where the counterexample also obeys Lemma 2.3 of [5].

In Theorem 2.5 of [5], we use such a counterexample to weak inclusion subtlety over 2 to construct a transitive set where the only proper inclusions among elements have left side $\emptyset$ or $\{\emptyset\}$. We want to carefully perform this construction here in Newcomp + Ext.

We define $D(S, f)$ if and only if

i) S is an x-system for some necessarily unique x;

ii) f is a univalent set of ordered pairs (function) with domain x;

iii) for all $y \in x$, $f(y) = \{f(z) : S(y, z)\}$.

We will need the following technical modification. $D(S, f, a)$ if and only if

i) S is an x-system for some necessarily unique x;

ii) $a \in x$;

iii) f is a univalent set of ordered pairs (function) with domain $\{b : \leq(x)(b, a)\}$;

iv) $\leq(x)(y, a) \rightarrow f(y) = \{f(z) : S(y, z)\}$.

Lemma 34. *Fix $D(S, f, a)$ and $D(S, g, b)$, where S is an x-system. f, g agree on their common domain. If $c \in x$ then the restriction h of f to $\{d : \leq(x)(d, c)\}$ has $D(S, h, c)$. f is one-one in the sense that for all $y, z \in \mathrm{dom}(f)$, if $<(x)(y, z)$ then $f(y) \neq f(z)$. For all $y, z \in \mathrm{dom}(f)$, $f(y) = f(z) \longleftrightarrow \equiv(y, z)$. The range of f is a virtual transitive set. For all $y, z \in \mathrm{dom}(f)$, if $f(y) \subseteq f(z)$ then $S_y \subseteq S_z$ and $\leq(x)(y, z)$. All proper inclusions among elements of the range of f have left side $\emptyset$ or $\{\emptyset\}$.*

Proof. Let S be an x-system. Suppose $f \neq g$, and let y be $<(x)$ minimal such that $f(x) \neq g(x)$. But $f(y) = \{f(z) : <(x)(z, y)\} = \{g(z) : <(x)(z, y)\} = g(y)$.

For the second claim, the restriction h is a subset of f, and so exists. $D(S, h, c)$ is immediate by inspection.

For the third claim, let $y, z \in \mathrm{dom}(f)$ be such that $<(x)(y, z)$, $f(y) = f(z)$, where y is chosen to be $<(x)$ minimal. By the strict sup condition, let $S(z, w)$, $\leq(y, w)$. Then $f(w) \in f(z)$, and so $f(w) \in f(y)$. Let $S(y, u)$, $f(w) = f(u)$. Then $<(x)(u, y)$, $\leq(y, w)$, and so the minimality of y is violated.

The fourth claim follows immediately from the third claim.

The fifth claim is immediate from clause iv).

For the sixth claim, assume $\{f(u) : S(y, u)\} \subseteq \{f(u) : S(z, u)\}$. Since f is appropriately one-one, we have $\{u : S(y, u)\} \subseteq \{u : S(z, u)\}$. By the strict sup condition, $\leq(x)(y, z)$.

For the seventh claim, suppose $f(y) \subsetneq f(z)$. By the sixth claim, $<(x)(y, z)$ and $S_y \subsetneq S_z$. By clause v) in the definition of x-system, we have $y = \emptyset$ or $y = \{\emptyset\}$. Hence $f(y) = \emptyset$ or $f(y) = \{\emptyset\}$. □

Let S be an x-system. We use S^* for the virtual set of all sets that occur in the range of some f with $(\exists a)(D(S, f, a))$.

Lemma 35. *Let S be an x-system. Then S^* is transitive and all proper inclusions among elements of S^* have left side $\emptyset$ or $\{\emptyset\}$. $x \cup S^*$ exists.*

Proof. The transitivity of S^* follows from the transitivity of the range of each relevant f. Let $u, v \in S^*$, where u, v are in the ranges of f, g, and $D(S, f, a)$, $D(S, g, b)$. By the comparability of a, b under $\leq(x)$ and Lemma 34, we have $f \subseteq g$ or $g \subseteq f$. Hence any proper inclusion among elements of S^* is a proper inclusion among some relevant h. So by the last claim of Lemma 34, the proper inclusion must have left side $\emptyset$ or $\{\emptyset\}$.

For the second claim, note that every expansion in $x \cup S^*$ meets $x \cup \{\emptyset\} \cup \{\{\emptyset\}\}$. The latter forms a set by Lemma 22. Hence $x \cup S^*$ exists. □

According to Lemma 24, $(x \cup S^*) \cdot (x \cup S^*)$ exists, and in fact we can simulate all Cartesian powers of $x \cup S^*$, and hence also simulate full second logic over $(x \cup S^*, \in)$.

Lemma 36. *Suppose S is an x-system, $b \in x$, and for all a with $<(x)(a, b)$, there exists f such that $D(S, f, a)$. Then there exists g such that $D(S, g, b)$.*

Proof. Let S, x, b be as given. By Lemma 34, for each a with $<(x)(a, b)$, there is a unique f_a such that $D(S, f_a, a)$, and furthermore, these f_a agree on their common domains. Clearly b is either the minimum of $<(x)$, a successor in $<(x)$, or a limit in $<(x)$. If b is the minimum of $<(x)$ then $b = \emptyset$ by Lemma 31, in which case set $g = \{\langle \emptyset, \emptyset \rangle\}$.

Assume b is the successor of c in $<(x)$. We need to extend the function g_c to be defined on $\{d : \equiv(x)(d, b)\}$ by $g(d) = \{f(z) : S(b, z)\}$. Note that this extension is a binary relation on the transitive virtual set $x \cup S^* \cup \{\{f(z) : S(b, z)\}\}$. By Lemma 35, $x \cup S^*$ exists, the triple union is transitive, and all expansions in the triple union meet $x \cup S^* \cup \{\emptyset, \{\emptyset\}\}$. Hence the triple union exists. Since the desired extension of the function g_c to g_b is a virtual binary relation on the triple union, it exists.

Assume b is a limit in $<(x)$. Let f be the virtual set which is the union of all f_a, $<(x)(a, b)$. This is obviously a binary relation on $x \cup S^*$, and therefore it exists. By Lemmas 34 and 35, this union is itself a function g with domain $\{a : <(x)(a, b)\}$, given by $g(a) = \{g(z) : S(b, z)\}$. We need to extend this function g to be defined on $\{d : \equiv(x)(d, b)\}$ by $g(d) = \{f(z) : S(b, z)\}$. We argue exactly as in the previous paragraph. □

Lemma 37. *Suppose S is an x-system. For all $b \in x$, there exists f such that $D(S, f, b)$. There exists g such that $D(S, g)$.*

Proof. Let $b \in x$ be $<(x)$ minimal such that there is no f with $D(S, f, b)$. Then the hypotheses of Lemma 36 hold, and we obtain a contradiction. For the second claim, the case $x = \emptyset$ is trivial. If there is a $<(x)$ greatest element b of x then any f with $D(S, f, b)$ has $D(S, f)$. Finally suppose x is nonempty and there is no $<(x)$ greatest element of x. By Lemma 35, $x \cup S^*$ exists. Thus the desired function is a binary

relation on $x \cup S^*$. Therefore we can take g to be the union of the f such that for some b, $D(S, f, b)$. As in the proof of Lemma 36, $D(S, g)$. □

Let S be an x-system. We define $S\#$ to be the unique (set) function such that $D(S, S\#)$, according to Lemma 37. Note that $S\# : x \to S^*$, $S\#$ is onto, and $x \cup S^*$ is a transitive set. In addition, for all $b \in x$, $S\#(b) = \{S\#(z) : S(b, z)\}$, and by Lemma 34, $S\#(b) = S\#(c) \longleftrightarrow \equiv(x)(b, c)$. We refer to this equivalence as the one-one property of $S\#$.

Lemma 38. *Let S be an x-system. $W(S^*)$. For all $a, b \in x$, $\leq(x)(a, b)$ if and only if $\leq(S^*)(S\#(a), S\#(b))$. $\equiv(S^*)$ is the identity relation on S^*. $<(S^*)$ is a strict well ordering.*

Proof. By Lemma 35, S^* is transitive. To prove well foundedness, let y be a nonempty subset of S^*. Let $z = \{w \in x : S\#(w) \in y\}$. Let w be a $<(x)$ minimal element of z. We claim that $S\#(w)$ is an epsilon minimal element of y. To see this, let $u \in S\#(w) \cap y$. Let $u = S\#(w') \in y$, $S(w, w')$. Then $w' \in z$, $<(x)(w', w)$, contradicting the $<(x)$ minimality of w. Hence $W(S^*)$.

The second claim now makes sense since $W(S^*)$. Recall the definition of $\leq(S^*)$ as the unique relation satisfying clauses i)–iii) given just before Lemma 25. To establish the second claim, we have only to show that the binary relation R on S^* given by $R(S\#(a), S\#(b)) \longleftrightarrow \leq(x)(a, b)$ obeys conditions i)–iii) with $x = y = S^*$. The only clause of substance is iii).

Let $z, w \in S^*$. We must verify that $R(z, w) \longleftrightarrow (\forall a \in z)(\exists b \in w)(R(a, b))$. In other words, let $a, b \in x$. We must verify that $R(S\#(a), S\#(b)) \longleftrightarrow (\forall u \in S\#(a))(\exists v \in S\#(b))(R(u, v))$.

The left side is equivalent to $\leq(x)(a, b)$. The right side is equivalent to $(\forall c | S(a, c))(\exists d | S(b, d))(\leq(x)(c, d))$. Suppose $<(x)(a, b)$. This is true because $S(a, c) \to <(x)(c, a)$ and the strict sup condition for x-systems. Suppose $\equiv(x)(a, b)$. This is true by setting $d = c$. For the inverse, suppose $<(x)(b, a)$. Choose c such that $S(a, c)$, $\leq(x)(b, c)$. Any d with $S(b, d)$ will have $<(x)(d, b)$, and therefore not $\leq(x)(c, d)$.

From the third claim, using Lemma 34, we have $\equiv(S^*)(S\#(a), S\#(b))$ if and only if $\equiv(x)(a, b)$ if and only if $S\#(a) = S\#(b)$. By $W(S^*)$ and Lemma 33, $<(S^*)$ is well founded and connected, and so by the third claim, $<(S^*)$ is a strict well ordering. □

The S^* are our best approximations to genuine ordinals in Newcomp + Ext. $<(S^*)$ is a strict well ordering. Of course, S^* is not normally epsilon connected.

Lemma 39. *Let S, S' be two x-systems. For all $y, z \in x$, $S\#(y) = S'\#(z) \to \equiv(x)(y, z)$. $S = S' \longleftrightarrow S^* = S'^*$.*

Proof. Let S, S' be as given. For the first claim, let y, z be a counterexample, with y chosen to be $<(x)$ minimal. We first assume that $>(x)(y, z)$. Now $S\#(y) = S'\#(z)$.

By the strict sup condition on x-systems, let $S(y, w)$, $\geq(x)(w, z)$. Then $S\#(w) \in S\#(y)$, and so $S\#(w) \in S\#(z)$. Let $S\#(w) = S'\#(u)$, $S'(z, u)$. Clearly $\neg \equiv(w, u)$. This contradicts the minimality of z.

Now assume $<(x)(y, z)$. By the strict sup condition on x-systems, let $S'(z, b)$, $\geq(x)(b, y)$. Then $S'\#(b) \in S'\#(z) = S\#(y)$. Let $S'\#(b) = S\#(c)$, $S(y, c)$. Clearly $\neg \equiv(b, c)$, and so we have a violation of the minimality of z. This establishes the first claim.

For the second claim, the forward direction is obvious. Now let $S^* = S'^*$. Let y, z be such that $S(y, z) \longleftrightarrow \neg S(y, z)$, where y is $<(x)$ minimal. By symmetry, we can assume $S(y, z)$, $\neg S(y, z)$.

An obvious transfinite induction (or minimal element) argument, shows that $S\#$ and $S'\#$ agree below y in the sense of $<(x)$.

We now show that $S\#(y) \neq S'\#(y)$. Suppose $\{S\#(z) : S(y, z)\} = \{S'\#(z) : S'(y, z)\}$. Then $S\#(z) \in \{S'\#(z) : S'(y, z)\}$. Let $S\#(z) = S'\#(w)$, $S'(y, w)$. Note that $\neg \equiv(w, z)$. By the previous paragraph, $S'\#(w) = S\#(w) = S\#(z)$. This contradicts the one-one property of $S\#$.

By the one-one property of $S\#$, $S'\#$, we see that $S\#(y)$ is not equaled to any $S'\#(z)$, $\equiv(x)(y, z)$. By the first claim, $S\#(y)$ is not equaled to any $S'\#(z)$, $\neg \equiv(x)(y, z)$. Hence $S\#(y)$ is an element of S^* that is not an element of S'^*. □

Let S be an x-system and S' be a y-system. We say that S, S' are equivalent if and only if $S^* = S'^*$.

We say that R is an isomorphism relation from $\leq(x)$ onto $\leq(y)$ if and only if the following holds. We do not require that R exists; i.e., R may only be a virtual relation. Since we have ordered pairs, R can always be viewed as a virtual set.

i) the domain of R is x and the range of R is y;

ii) if $R(a, b)$ and $R(c, d)$ then $\leq(x)(a, c) \longleftrightarrow \leq(y)(b, d)$.

We say that $\leq(x)$ and $\leq(y)$ are isomorphic if and only if there is an isomorphism relation from $\leq(x)$ onto $\leq(y)$.

Lemma 40. *Let x, y be given. Suppose some x-system is equivalent to some y-system. Then $\leq(x)$ and $\leq(y)$ are isomorphic. If $\leq(x)$ and $\leq(y)$ are isomorphic, then every x-system is equivalent to some y-system and vice versa.*

Proof. Before we begin, we make some comments about the formalization of Lemma 40. The problem arises because the definition of isomorphic $\leq(x)$ involves virtual relations. For the second claim, this just means that we are making infinitely many assertions. For the first claim, we mean that the virtual isomorphism can be given uniformly in x, y, and the x-system and the y-system; i.e., by a single specific virtual relation with parameters, as will be clear in the construction below.

Let S be an x-system, S' be a y-system, $S^* = S'^*$. We define the virtual relation R from x to y as follows. Let $a \in x$ and $b \in y$. Take $R(a, b) \longleftrightarrow S\#(a) = S'\#(b)$. We claim that R is an isomorphism relation from $\leq(x)$ onto $\leq(y)$.

To see that the domain of R is x, let $a \in x$. Then $S\#(a) \in S^* = S'^*$. Let $b \in y$, $S'\#(b) = S\#(a)$. Then $R(a, b)$. Analogously, the range of R is y.

Now let $R(a, b)$, $R(c, d)$. By Lemma 38, $\leq(x)(a, c) \longleftrightarrow \leq(S^*)(S\#(a), S\#(c))$, and $\leq(y)(b, d) \longleftrightarrow \leq(S'^*)(S'\#(b), S'\#(d))$. Hence $\leq(y)(b, d) \longleftrightarrow \leq(S^*)(S\#(a), S\#(c))$. So $\leq(x)(a, c) \longleftrightarrow (y)(b, d)$.

Now let R be a virtual isomorphism relation from $\leq(x)$ onto $\leq(y)$. We now show that every x-system is equivalent to some y-system. Let S be an x-system. Define S' to be the binary relation on y given by $S'(z, w)$ if and only if there exists $a, b \in x$ such that $R(a, z)$, $R(b, w)$, and $S(a, b)$. Note that S' exists since it is a virtual subset of the set $y \cdot y$.

We have to verify that $S^* = S'^*$. I.e., that $S\#$ and $S'\#$ have the same ranges. It suffices to show that $R(a, b) \rightarrow S\#(a) = S'\#(b)$. Let $a, b \in x$ be a counterexample with a chosen to be $<(x)$ minimal. We can find such a, b by using separation on x, despite the fact that R is only virtual. Observe that $S\#(a) = \{S\#(c) : S(a, c)\}$, $S'\#(b) = \{S'\#(d) : S'(b, d)\}$. We have $R(a, b)$, $S\#(a) \neq S'\#(b)$.

We first show that $S\#(a) \subseteq S'\#(b)$. We will then show that $S'\#(b) \subseteq S\#(a)$, and so $S\#(a) = S'\#(b)$. This contradicts the choice of a, b.

Let $S(a, c)$. We show that $S\#(c) \in \{S'\#(d) : S'(b, d)\} = S'\#(b)$. Since $R(a, b)$, we let $d \in y$ be such that $S'(b, d)$ and $R(c, d)$. By the minimality of a, we have $S\#(c) = S'\#(d)$. Hence $S\#(c) \in \{S'\#(d) : S'(b, d)\} = S'\#(b)$. This establishes $S\#(a) \subseteq S'\#(b)$.

To establish $S'\#(b) \subseteq S\#(a)$, let $S'(b, d)$. We show that $S'\#(d) \in \{S\#(c) : S(a, c)\}$. Since $R(a, b)$, we let $c \in x$ be such that $S(a, c)$ and $R(c, d)$. By the minimality of a, we have $S'\#(d) = S\#(c)$. Hence $S'\#(d) \in \{S\#(c) : S(a, c)\} = S\#(a)$. This establishes $S'\#(b) \subseteq S\#(a)$. □

We refer to the conclusion of the second claim of Lemma 40 as "x, y have equivalent systems".

Let $W(x)$. The initial segments of x are the subsets of x closed under $\leq(x)$. The proper initial segments are written $x_{<a}$, $a \in x$. The initial segments of an x-system S also has the obvious meaning, and the proper initial segments are written $S_{<a}$, $a \in x$. We also use the notation $x_{\leq a}$, $S_{\leq a}$, with the obvious meaning. And we also use the notation $S\#_{\leq a}$, $S\#_{<a}$, again with the obvious meaning.

Lemma 41. *Suppose that there exists an x-system. It is not the case that x and one of its proper initial segments have equivalent systems.*

Proof. Suppose x and $x_{<a}$ have equivalent systems. By Lemma 40, $\leq(x)$ and $\leq(x_{<a})$ are isomorphic. This violates the well foundedness of x, using separation on x, even if the isomorphism is virtual. □

Lemma 42. *Let x, y have systems, and assume that for every proper initial segment of x there is a proper initial segment of y with equivalent systems, and vice versa. Then x, y have equivalent systems.*

Proof. Let x, y be as given. Suppose $x_{<a}$ and $y_{<b}$ have equivalent systems. Since x, y have systems, so do $x_{<a}$ and $y_{<b}$. By Lemma 41, $y_{<b}$ is the unique initial segment of y such that $x_{<a}$ and $y_{<b}$ have equivalent systems.

Define the virtual relation $R(a, b)$ if and only if $x_{<a}$ and $y_{<b}$ have equivalent systems. We claim that R is a virtual isomorphism from $\leq(x)$ onto $\leq(y)$. By hypothesis, the domain of R is x and the range of R is y. Suppose $R(a, b)$ and $R(c, d)$. Then $x_{<a}$ and $y_{<b}$ have equivalent systems. Also $x_{<c}$ and $y_{<d}$ have equivalent systems. By Lemma 40, we obtain a virtual isomorphism from $x_{<a}$ onto $y_{<b}$, and a virtual isomorphism from $x_{<c}$ onto $y_{<d}$. Suppose $\leq(x)(a, c)$, and $<(x)(d, b)$. Then we get a virtual isomorphism from $x_{<a}$ onto an initial segment of $y_{<d}$, and hence onto a proper initial segment of $y_{<b}$. But this gives a virtual isomorphism from $y_{<b}$ onto a proper initial segment of $y_{<b}$, which is a virtual subset of the set $y \cdot y$. Hence using separation, we have an actual isomorphism from $y_{<b}$ onto a proper initial segment of $y_{<b}$, which contradicts the well foundedness of $<(y)$. The other direction is symmetric.

We have thus verified that R is a virtual isomorphism from $\leq(x)$ onto $\leq(y)$. By Lemma 40, x, y have equivalent systems. □

Lemma 43. *Let x, y have systems. Either x and some proper initial segment of y have equivalent systems, or y and some proper initial segment of x have equivalent systems, or x, y have equivalent systems. The cases are mutually exclusive, and the choice of proper initial segments is unique.*

Proof. Let x, y have systems. Let $B = \{a \in x : (\exists b \in y)(x_{<a}$ and $y_{<b}$ have equivalent systems)$\}$. We claim that $(\leq(x)(c, a) \wedge a \in B) \rightarrow c \in B$. To see this, assume $x_{<a}$ and $y_{<b}$ have equivalent systems. By Lemma 40, we obtain a virtual isomorphism from $x_{<a}$ onto $y_{<b}$. This induces a virtual isomorphism from $x_{<c}$ onto some $y_{<d}$. By Lemma 40, we see that $x_{<c}$ and $y_{<d}$ have equivalent systems.

Note that for $a \in B$, the associated $b \in y$ is unique up to $\equiv(y)$. We let C be the set of all these associated $b \in y$ that are used in B, closing up under $\equiv(y)$. As in the previous paragraph, C is also closed downward under $\leq(y)$.

By the well foundedness of $<(x)$, either $B = x$ or B is a proper initial segment of x. Also either $C = y$ or C is a proper initial segment of y.

No matter which of these four cases holds, we can apply Lemma 42. We obtain that B, C have equivalent systems. This establishes the first claim, provided we can rule out the fourth case where B is a proper initial segment of x and C is a proper initial segment of y. Let $B = x_{<a}$, and $C = y_{<b}$. Then $a \in B$, and so $<(x)(a, a)$, which is impossible.

For the second claim, if more than one case holds, or if the choice of proper initial segments is not unique, then we obtain that x and a proper initial segment of x have equivalent systems, or y and a proper initial segment of y have equivalent systems. This violates Lemma 41. □

Suppose x, y have systems. By Lemma 43, either x and some unique proper initial segment of y have equivalent systems, or y and some unique proper initial segment of

x have equivalent systems, or x, y have equivalent systems, and these three cases are mutually exclusive. By Lemma 42, we obtain a (rather simple) virtual isomorphism relation corresponding to these three cases. The relevant virtual isomorphism relations are defined uniformly in x, y. We call it the virtual comparison relation for x, y.

In light of the previous paragraph, we make the following definition, assuming x, y have systems. COMP$(x, y)(a, b)$ if and only if the virtual comparison relation for x, y holds at a, b.

We now construct an x such that $\leq(x)$ has "order type ω". The idea is to construct $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots\}$.

We say that x is simple if and only if

i) $W(x)$;

ii) every nonempty element of x is the singleton of an element of x.

We say that x is very simple if and only if x is simple and x has an epsilon maximal element.

Lemma 44. *Let x be simple and y be an epsilon maximal element of x. x is the least transitive set containing y as an element. The singleton of any element other than y is an element. No simple set has more than one epsilon maximal element.*

Proof. Let x, y be as given. Let z be a transitive set with $y \in z$. We claim that $x \subseteq z$. First observe that $x \cap z$ is nonempty, transitive and well founded, and so $\emptyset \in x \cap z$. Now let b be an epsilon minimal element of x not in z. Then $b \neq \emptyset$. Write $b = \{c\}$, $c \in x$. Since z is transitive, $c \in z$. This contradicts the minimality of b.

For the second claim, let z be the set of all elements of x whose singleton lies in x, together with the element y. To see that z is transitive, let $a \in b \in z$. Then b is a nonempty element of x, and so $b = \{a\}$. Therefore $a \in z$. By the first claim, $x \subseteq z$.

The third claim follows immediately from the second. □

Lemma 45. *The very simple sets are comparable under inclusion.*

Proof. Let x, y be very simple. Assume x is not a subset of y. Let b be an epsilon least element of x that is not in y. If $b = \emptyset$ then y is empty, in which case $y \subseteq x$. So we assume $b \neq \emptyset$. Let $b = \{c\}$, $c \in x$. Then $c \in y$. By Lemma 44, if c is not the epsilon maximal element of y, then $b = \{c\} \in y$. Therefore c is the epsilon maximal element of y. By Lemma 44, y is the least transitive set containing c as an element. But x is a transitive set containing c as an element. Hence $y \subseteq x$. □

Lemma 46. *In a very simple set, every proper inclusion between elements has left side $\emptyset$. If x is a very simple set with epsilon maximal element y, then $x \cup \{\{y\}\}$ is very simple.*

Proof. Let $a, b \in x$, $a \subsetneq b$. Then $b \neq \emptyset$, and so write $b = \{c\}$, $c \in x$. Then $a = \emptyset$.

Let x be very simple with epsilon maximal element y. Clearly $x \cup \{\{y\}\}$ is transitive and well founded. Let $b \in x \cup \{\{y\}\}$, $b \neq \emptyset$. If $b \in x$ then $b = \{c\}$ for some $c \in x \cup \{\{y\}\}$. If $b = \{y\}$ then just note that $y \in x \cup \{\{y\}\}$.

It remains to show that $\{y\}$ is an epsilon maximal element of $x \cup \{\{y\}\}$. Now $\{y\} \in \{y\}$ is impossible since then $\{y\} = y$, and so $y \in y$, violating the well foundedness of x. It remains to show that no element of x includes $\{y\}$ as an element. Let $\{y\} \in b \in x$. Then $\{y\} \in x$, violating that y is the epsilon maximal element of x. □

Lemma 47. *There is a unique simple set with no epsilon maximal element. It is the union of all very simple sets. Every proper inclusion among elements of this set has left side* $\emptyset$.

Proof. Let A be the virtual union of all very simple sets. Let $a \subsetneq b$, $a, b \in A$. Let $a \in x$, $b \in y$, where x, y are very simple. By Lemma 45, we can assume $a, b \in x$. By Lemma 45, $a = \emptyset$. Therefore all expansions in A meet $\{\emptyset\}$. Hence A exists.

To see that A is simple, we first check that $W(A)$. Clearly A is transitive since it is the union of transitive sets. To see that A is well founded, let z be a nonempty subset of A. Let x be very simple, where x, z have an element in common. Let b be an epsilon minimal element of $x \cap z$. We claim that b is an epsilon minimal element of z. This follows from the transitivity of x.

To finish the argument that A is simple, let b be a nonempty element of A. Then $b \in x$ for some very simple x. Hence $b = \{c\}$ for some $c \in x$. Hence $b = \{c\}$ for some $c \in A$.

We now show that A has no epsilon maximal element. Let $b \in x$, where x is very simple, with epsilon maximal element y. By Lemma 44, $b = y$ or $\{b\} \in x$. If $\{b\} \in x$ then $\{b\} \in A$ and b is not an epsilon maximal element of A. Now assume $b = y$. By Lemma 46, $x \cup \{\{y\}\}$ is very simple. Hence again $\{b\} \in A$, and so b is not an epsilon maximal element of A.

We now prove that A is the unique simple set with no epsilon maximal element. It suffices to prove that for any two simple sets y, z with no epsilon maximal elements, we have $y \subseteq z$. Suppose this is false, and let b be an epsilon minimal element of y that is not in z. Then $b \neq \emptyset$, since z is nonempty and transitive. Let $b = \{c\}$, $c \in y$. Then $c \in z$. Since c is not epsilon maximal in z, let $c \in d \in z$. Then $d = \{e\}$ for some $e \in z$. Hence $c = e$ and $d = \{c\} = b \in z$.

By the first paragraph, every proper inclusion among elements of A has left side $\emptyset$. □

Lemma 48. *The unique simple set with no epsilon maximal element is the least set containing* $\emptyset$ *as an element, and closed under the singleton operation.*

Proof. This unique set is $A =$ the union of all very simple sets, according to Lemma 47. Obviously $\emptyset \in A$, because $\{\emptyset\}$ is very simple. Also by Lemmas 44 and 46, A is closed under the singleton operation.

Now let y be any transitive set containing $\emptyset$ as an element, and closed under the singleton operation. Suppose A is not a subset of y. Let x be an epsilon minimal element of A that is not in y. Clearly $x \neq \emptyset$ since $\emptyset \in y$. Since A is simple, let $x = \{b\}$, $b \in A$. Then $b \in y$, and so $x = \{b\} \in y$. □

We use the notation $\omega\hat{}$ for the set in Lemma 48, which is the union A of all very simple sets.

Note that $\omega\hat{}$ with $\emptyset$ as 0 and the singleton operation as successor forms a successor structure with second order induction in the following standard sense. The successor of any element is not 0. Any two elements with the same successor are equal. Every set that includes 0 as an element, and is closed under successor, contains the whole structure as a subset.

This means that we have full second order arithmetic at our disposal for use in the rest of the section.

Lemma 49. *$W(\omega\hat{})$. $\leq(\omega\hat{})$ is a well ordering with no greatest element and no limit point.*

Proof. $W(\omega\hat{})$ since $\omega\hat{}$ is simple. Hence $\leq(\omega\hat{})$ is well founded.

To obtain that $\leq(\omega\hat{})$ is a well ordering, it suffices to show that $\equiv(\omega\hat{})(x, y) \rightarrow x = y$. Let x, y be a counterexample, where x is epsilon minimal. If $x = \emptyset$ then from $\leq(\omega\hat{})(y, \emptyset)$ we obtain $y = \emptyset$. So $x \neq \emptyset$. Let $x = \{b\}$. We argue similarly that $y \neq \emptyset$, and so let $y = \{c\}$. Since $x \neq y$, we have $b \neq c$. Since $\equiv(\omega\hat{})(x, y)$, we have $\leq(\omega\hat{})(x, y)$, $\leq(\omega\hat{})(x, y)$. By the definition of $\leq(\omega\hat{})$, we have $\leq(\omega\hat{})(b, c)$, $\leq(\omega\hat{})(c, b)$, and so $\equiv(\omega\hat{})(b, c)$. This contradicts the minimality of x.

That $\leq(\omega\hat{})$ has no greatest element follows from the fact that $\omega\hat{}$ has no epsilon maximal element, the first claim, and Lemma 31. □

$\omega\hat{}$ certainly gives us full second order arithmetic, but how are we going to use it to interact with the binary relations on any x with $W(x)$? This is answered by the following.

Lemma 50. *If $W(x)$ then $x \cup \omega\hat{}$ exists. In fact, $W(x \cup \omega\hat{})$.*

Proof. By Lemma 47, every proper inclusion between elements of $\omega\hat{}$ has left side $\emptyset$. Therefore all expansions in $x \cup \omega\hat{}$ meet $x \cup \{\emptyset\}$. Hence $x \cup \omega\hat{}$ exists. The second claim follows easily. □

This means that what is essentially the Cartesian powers of $x \cup \omega\hat{}$, and therefore separation on them, are available.

Lemma 51. *Suppose there is an x-system. Then there is an $x \cup \{x\}$-system.*

Proof. Let S be an x-system. Obviously $W(x \cup \{x\})$ since transitivity and well foundedness are preserved. In Lemma 2.3 of [5], a construction is given that in our contexts

amounts to modifying S to another x-system with the property that every cross section contains an element at the bottom level (i.e., 0), or an element at an odd level (i.e., odd as in odd ordinal level). These levels refer to levels in $\leq(x)$. Bringing this construction into this context requires ordinal arithmetic, which is fully available here. An $x \cup \{x\}$-system is formed by augmenting the x-system with a top level cross section which is the set of all elements of x that lie at a nonzero even level in $\leq(x)$. □

Note that $\leq(\omega\hat{})$ has no greatest element. We want to extend Lemma 57 by constructing $x \cup \{x, \{x\}, \{\{x\}\}, \dots\}$, with this same property.

Lemma 52. *Assume that there is an x-system. There exists a set z such that $x \in z$ and z is closed under singletons. There is a least such z. There is an $x \cup z$-system. $\leq(x \cup z)$ has no greatest element. $\leq(x)$ is a proper initial segment of $\leq(x \cup z)$.*

Proof. We use $\omega\hat{}$ for the construction. By induction, for each $n \in \omega\hat{}$, there is a unique function f_n with domain $\{i : \leq(\omega\hat{})(i, n)\}$ such that $f(\emptyset) = x$ and each $f(\{i\}) = \{f(i)\}$. Using $W(x)$, each f_n is one-one. The union of the ranges of the f_n is a virtual set B containing x as an element and closed under singletons. Note that every expansion within the virtual set $x \cup B$ meets $x \cup \{x\}$. Hence $x \cup B$ exists. Hence B exists. Clearly B is contained in any set z such that $x \in z$ and z is closed under singletons. Therefore B serves as the z.

We now verify $W(x \cup z)$. By induction, every element of $B = z$ is either x or the singleton of an element of z. Hence $x \cup z$ is transitive. For well foundedness, if the nonempty subset of $x \cup z$ meets x then take an epsilon minimal element in x. Show that this is epsilon minimal by induction on the construction of z. Otherwise, also use induction in the construction of z.

To see that there is an $x \cup z$-system, start with the $x \cup \{x\}$-system provided by Lemma 51. Extend it to the rest of $x \cup z$ by taking the cross section at each $b \in z \backslash \{x\}$ to consist of just the unique element of b. Obviously $\leq(x \cup z)$ has no greatest element since $x \cup z$ has no epsilon maximal element. Also, $\leq(x)$ is the proper initial segment of $\leq(x \cup z)$ determined by the point x. □

We now wish to build the constructible hierarchy on every x with $W(x)$. We run into trouble if we wish to compare the constructible hierarchy on x with the constructible hierarchy on y, where $W(x)$, $W(y)$. There is where we will need that x, y have systems; i.e., there is an x-system and there is a y-system.

Rather than reinvent the wheel, we will use a fairly strong version of a standard sentence σ in $L(\in)$ used in standard treatments of the constructible hierarchy in set theory with the property that the well founded models of σ are exactly the structures which, when factored by the equivalence relation of extensional equality, are isomorphic to some $(L(\lambda), \in)$, where λ is a limit ordinal. This is normally done in $L(\in, =)$ in connection with the verification of the generalized continuum hypothesis in L, but here we stay within $L(\in)$.

We wish to be more explicit about σ for two reasons. One is for the sake of expositional completeness, and the other is because we are going to use σ formally, especially in Lemma 59.

We take σ to be the conjunction of the following sentences in $L(\in)$. We do not make much effort to be economical.

i) extensionality, pairing, union, every set has a transitive closure, every nonempty set has an epsilon least element, every set has a cumulative rank function, every set whose cumulative rank function is onto a finite ordinal has a power set, there is no greatest ordinal;

ii) if there is a limit ordinal then the satisfaction relation of every set under epsilon exists;

iii) if there is a limit ordinal then for all ordinals α, there is a function on α which follows the usual definition by transfinite recursion of the constructible hierarchy;

iv) if there is a limit ordinal then every set lies somewhere in the constructible hierarchy defined in iii);

v) Δ_0 separation.

Here an ordinal is an epsilon connected transitive set. A finite ordinal is an ordinal which is not a limit ordinal and no element is a limit ordinal. A cumulative rank function on x is an ordinal valued function f with domain x such that each $f(y)$ is the strict sup of the $f(z)$, $z \in y$. It is well known how to finitely axiomatize Δ_0 separation in the presence of extensionality, pairing, and union. In ii), the satisfaction relation is defined using finite sequences from the set for the assignments, where a finite sequence is taken to be a function from an element of the least limit ordinal into the set.

Lemma 53. *Let $W(x)$, $x \neq \emptyset$, and $\leq(x)$ have no greatest element. There exists $A \subseteq x \cdot x$ and binary relation $R \subseteq A \cdot A$ such that (A, R) satisfies σ, R is well founded, and the relative level in the internal constructible hierarchy of an ordered pair in A is given by its first coordinates position in $\leq(x)$, and every element of x is the first coordinate of an element of A. I.e., for all (a, b), $(c, d) \in A$, (A, R) satisfies "(a, b) occurs at an earlier stage in the constructible hierarchy (the $L(\alpha)$'s) than (c, d)" if and only if $<(x)(a, c)$. Furthermore, equivalence (having the same elements) in the sense of (A, R) is the same as the equivalence relation E on A given by $(a, b)\ E\ (c, d) \longleftrightarrow (\equiv(x)(a, c) \wedge \equiv(x)(b, d))$.*

Proof. By Lemma 50, $x \cup \omega\hat{}$ exists. So we can simulate any of its Cartesian powers and use separation on them. We can treat $\leq(x)$ as a well ordering, provided that for each $b \in x$, we carry along all c with $\equiv(x)(b, c)$. We can build a coded version of the constructible hierarchy up through $\leq(x)$ (in fact well beyond that point) as a

binary relation on an appropriate subset of $x \cdot x$. There is no difficulty assembling this construction so as to be a binary relation on an appropriate subset of $x \cdot x$ where the first coordinate is used for the level of the constructible hierarchy. We use the $\omega\hat{}$ in $x \cup \omega\hat{}$ in order to code formulas and set up the needed satisfaction relations for the successor steps in the constructible hierarchy. One delicate point is that we need to use a function from the finite sequences from any given infinite initial segment of $\leq(x)$, mod $\equiv(x)$, into that same initial segment, mod $\equiv(x)$. Since we don't properly have the set of all such finite sequences in this environment, we use a suitable mapping from any given infinite initial segment of $\leq(x)$, cross $\omega\hat{}$, into that same initial segment, which serves as an appropriate finite sequence mechanism. Because we have what amounts to full second order logic on $(x \cup \omega\hat{}, \in)$ at our disposal, we can explicitly provide such a finite sequence mechanism with the help of some relevant ordinal arithmetic. □

The construction of (A, R) in Lemma 53 is done uniformly in x. So for $x \neq \emptyset$, $W(x)$, $\leq(x)$ with no greatest element, we write $L[x]$ for the (A, R), $A \subseteq x \cdot x$, $R \subseteq A \cdot A$, given by the proof of Lemma 53. We write $\mathrm{dom}(L[x])$ for the A. The idea of the notation is that this is the code for the initial segment of the constructible hierarchy along $\leq(x)$.

We use the obvious notions of initial segment and proper initial segment of $L[x]$, where we only cut off at limit stages. We also use the obvious notions of isomorphism relations between initial segments of $L[x]$ and initial segments of $L[y]$.

It is convenient to define $W'(x)$ if and only if $x \neq \emptyset$, $\leq(x)$ has no greatest element, and there is an x-system. Note that $W'(\omega\hat{})$. In particular, $R \subseteq \omega\hat{} \cdot \omega\hat{}$ given by $R(n, m) \longleftrightarrow n = \{m\}$ is an $\omega\hat{}$-system.

Recall the definition of the virtual relation $\mathrm{COMP}(x, y)$ after the proof of Lemma 42.

Lemma 54. *Let $W(x)$, $W(y)$. There is a virtual isomorphism relation from $L[x]$ onto a proper initial segment of $L[y]$, or a virtual isomorphism relation from $L[y]$ onto a proper initial segment of $L[x]$, or a virtual isomorphism relation from $L[x]$ onto $L[y]$. The choice of cases and proper initial segments, as well as the comparison of levels, is by* $\mathrm{COMP}(x, y)$. *Furthermore, at most one of the three possibilities can apply, the proper initial segment is unique, and the isomorphism is unique.*

Proof. We suppose that $\mathrm{COMP}(x, y)$ is an isomorphism relation from $\leq(x)$ onto $\leq(y)$. The other cases are handled analogously.

$\mathrm{COMP}(x, y)$ obviously provides the right way of matching levels, but does not give us the desired virtual isomorphism relation. For this, we have to use the systems. Let S be an x-system, S' a y-system, where S, S' are equivalent. I.e., $S^* = S'^*$. Since $L[x]$ respects the equivalence relation $\equiv(x)$, we can forward image $L[x]$ via $S\#$ to get $A^* \subseteq S^* \cdot S^*$ and $L^*[x] \subseteq A^* \cdot A^*$. Similarly, we forward image $L[y]$ via $S\#$ to get $B^* \subseteq S^* \cdot S^*$ and $L^*[y] \subseteq B^* \cdot B^*$. Both of these forward images also satisfy σ, are well founded, and the level in the internal constructible hierarchy of an ordered pair is given by its first coordinate's position in $\leq(S^*)$. It is not clear

that $L^*[x] = L^*[y]$. However, $S^* \cup \omega^\wedge$, its Cartesian powers, and full separation are available, and so we can prove that $L^*[x]$ and $L^*[y]$ are isomorphic, by an actual, not merely virtual, isomorphism. This yields the desired virtual isomorphism relation from $L[x]$ onto $L[y]$ by composition. The comparison of levels by $\mathrm{COMP}(x, y)$ is preserved.

The final claim is by obvious transfinite induction (or minimal element) arguments. Firstly, observe that any virtual isomorphism must preserve levels in the constructible hierarchy, since the levels are well founded (here we use only separation on x and y separately). The uniqueness of the isomorphism constitutes infinitely many statements. These are again proved by obvious transfinite induction (or minimal element) arguments. We do these arguments separately on x and on y, and do not need $x \cup y$. □

For x, y with $W'(x)$, $W'(y)$, we let $\mathrm{LCOMP}(x, y)$ be the unique virtual comparison isomorphism relation between $L[x]$ and $L[y]$ given by the proof of Lemma 54. Note that $\mathrm{LCOMP}(x, y)$ is given uniformly in x, y.

We are now prepared to construct the full constructible universe. The points will be pairs (x, b), where $W'(x)$ and $b \in \mathrm{dom}(L[x])$. The epsilon relation between points (x, b) and (y, c) holds if and only if $\mathrm{LCOMP}(x, y)(b)$ is satisfied in $L[y]$ to be an element of c. We caution the reader that $\mathrm{LCOMP}(x, y)$ is not a function, but rather an isomorphism relation. However, it is functional when we factor out by the equivalence relations $\equiv(x)$ and $\equiv(y)$.

We use L for the virtual set of these points, and $\in_L$ for this virtual epsilon relation on L. Thus our notation for the full constructible universe in this context is $(L, \in_L)$. Both coordinates are virtual.

Lemma 55. *$(L, \in_L)$ is well founded in the sense that every nonempty subset of L has an $\in_L$ minimal element.*

Proof. Let $z \subseteq L$ be nonempty. Let $(y, c) \in z$, L. By $W'(y)$, let (x, b) be chosen such that $\mathrm{LCOMP}(x, y)(b)$ exists and is of minimum level in the constructible hierarchy of $L[y]$.

We claim that (x, b) is an $\in_L$ minimal element of z. To see this, let $(w, d) \in_L (x, b)$, $(w, d) \in z$, L. Then $\mathrm{LCOMP}(w, x)(d)$ is satisfied in $L[x]$ to be an element of b, and therefore is of lower level than b in the constructible hierarchy of $L[x]$. Therefore $\mathrm{LCOMP}(w, y)(d)$ is of lower level than c in the constructible hierarchy of $L[y]$. This contradicts the minimality of (x, b). □

Let σ' be the following variant of σ: any set is an element of a well founded transitive set satisfying σ.

Lemma 56. *σ' logically implies σ.*

Proof. Assume σ'. Then any list of sets of standard integer length all lie in some well founded transitive set satisfying σ.

To derive extensionality, let x, y have the same elements, and let z be given. Let $x, y, z \in w$, where w is a transitive set satisfying σ. Then x, y have the same elements in the sense of σ, and hence $x \in z \longleftrightarrow y \in z$.

To derive foundation, let x be a nonempty set, and $x \in y$, y a transitive set satisfying σ. Then x has an epsilon minimal element in the sense of y. So x has an epsilon minimal element.

We leave pairing and union to the reader.

Let x be given, and let $x \in y$, where y is a well founded transitive set satisfying σ. Then according to y, x has a cumulative rank function. Hence it is easily seen that this cumulative rank function in the sense of y is a cumulative rank function.

To establish Δ_0 separation, put all the parameters in some well founded transitive set, and apply Δ_0 separation inside there.

Let x be a well founded transitive set satisfying σ. Then the cumulative rank function for x exists. It is easy to see that its range is a limit ordinal. From this we obtain the ordinal ω as the least limit ordinal. We can apply induction to ω, as long as the predicate forms a subset of ω. In particular, we have Δ_0 induction.

We claim that every x is an element of some transitive set y satisfying σ where $\omega \in y$. To see this, let $x \in y \in z$, where y, z are transitive sets satisfying σ. Then the cumulative rank function for y exists in z, and hence $\omega \in z$.

To derive that x has a transitive closure, let $x \in y$, where y is a transitive set satisfying σ. Then x has a transitive closure u, in the sense of y. Clearly $x \subseteq u$ and u is transitive. Now the characterization of the transitive closure as the set of terms in backwards epsilon chains is provable from σ. Hence this characterization holds in y. To verify that u is the actual transitive closure, we use Δ_0 induction.

Suppose that the cumulative rank function of x is onto a finite ordinal. We claim that all subsets of x lie in y. To see this, let $z \subseteq x$. We show that the intersection of z with the first i elements of x lies in y by Δ_0 induction on i, using the cumulative rank function of x.

We now show that the satisfaction relation for any $(x, \in)$ exists. Write $x \in y \in z$, where y, z are transitive sets satisfying σ. Then z satisfies that there is a limit ordinal. So z satisfies that $(x, \in)$ has a satisfaction relation, and therefore $(x, \in)$ has a satisfaction relation.

Let α be an ordinal, $\alpha \in x$, x a transitive set satisfying σ. The constructible hierarchy as a function defined by transfinite recursion on α exists in the sense of x. Therefore the constructible hierarchy as a function defined by transfinite recursion on α exists.

Let y be a transitive set satisfying σ, with $\omega \in y$. The internal constructible hierarchy in y exhausts y. As in the previous paragraph, the internal constructible hierarchy in y is an initial segment of the external constructible hierarchy. Since every x lies in some transitive set satisfying σ which also has the element ω, it follows that every set appears in the constructible hierarchy. □

Lemma 57. *$(L, \in_L)$ is a well founded model of σ' and hence of σ.*

Proof. By Lemmas 55 and 56, we have only to verify that σ' holds in $(L, \in_L)$. Let $W'(x)$. By Lemma 52, let $W'(y)$, where $\leq(x)$ is a proper initial segment of $\leq(y)$. Look at the constructible hierarchy internal to $L[y]$. Internally, one of the points will be an $L(\alpha)$ where, externally, the α is the length of $\leq(x)$. Let b be such a point. Then in $(L, \in_L)$, (y, b) will be satisfied to be a well founded transitive set satisfying σ. Also, every $(x, a) \in L$ will have $(x, a) \in_L (y, b)$. □

Recall the scheme SVWISUB, which we have only discussed in the context of ZFC. We wish to discuss it in the context of σ. It is formulated identically.

Let # be the following scheme, which is a weakening of SVWISUB.

#. If ψ defines a system $A_\alpha \subseteq \alpha$, for all ordinals α, where the strict sup of each A_α is α, then there exist $2 \leq \alpha < \beta$ such that $A_\alpha \subseteq A_\beta$.

Lemma 58. SVWISUB *is provable in* $\sigma + \#$.

Proof. By a straightforward adaptation of Lemma 2.3 of [5]. This requires only the development of some simple ordinal arithmetic in σ. □

Lemma 59. ZFC + $V = L$ + SVWISUB *is provable in* $\sigma + \# + \neg$SUB.

Proof. By Lemma 58, we already have SVWISUB. Also $V = L$ follows immediately from σ. We have only to obtain ZFC.

Note that Lemma 3 can be proved just from σ, and so there is no weakly inclusion subtle ordinal over 2.

By an obvious use of SVWISUB, we obtain the existence of a limit ordinal. Let ω be the least limit ordinal. Then it is easily verified that ω obeys the axiom of infinity.

Next we verify replacement on ω. This is a strong analog of "On is uncountable" for eventual use in the adaptation of the proof of Theorem 1.6 of [5].

Suppose $\alpha_0 < \alpha_1 < \cdots$ is unbounded, and definable. Pick a counterexample to the weak inclusion subtlety over 2 of α_0. Next pick a counterexample to the weak inclusion subtlety over 2 of α_1, and place the part of the assignment to ordinals $\geq \alpha_0$ on top. Continue in this way.

There is no problem making the construction in Lemma 2.3 of [5] in this context for $\lambda = \alpha_i$ and $\delta = 2$, uniformly in i. In fact, we can adjust the "$2x + 1$" so that in the sets assigned to the limit ordinals for $\lambda = \alpha_i$, the least integer present is always $2i + 1$. This guarantees that we have given a counterexample to SVWISUB. Hence we have replacement on ω.

Now we verify the analog of "λ is a cardinal". Specifically, that there is no definable map from On one-one into an ordinal. This is a straightforward adaptation of Lemmas 1.4 and 1.5 of [5].

We now wish to prove the analog of "λ is subtle" by adapting the proof of Lemma 1.6 of [5]. We first need the analog of Theorem 1.2 of [5], which is "On is inclusion subtle implies On is subtle". This is no problem. Here these notions are, as always, formulated on On through definable assignments. We already have

"On is an uncountable cardinal" by the previous paragraph and ω replacement. The adaptation of the proof of Theorem 1.6 of [5] is clear and shows that "On is subtle".

It is now easy to verify replacement. Suppose replacement fails on α. Then we obtain a definable mapping partially from α into ordinals which is unbounded. We can adjust this map so that the range, C, is closed and unbounded. We can then give an easy counterexample to the subtlety of On by assigning singletons to sufficiently large elements of C.

We now verify separation. Separation can be proved from σ together with replacement. It suffices to fix an ordinal α and find a limit ordinal $\lambda > \alpha$ such that $L(\lambda)$ is an elementary substructure of L with respect to n quantifier prenex formulas, for any standard integer n. It is clear what we mean by the first k stages $L(\alpha_0)$, $L(\alpha_1)$, $L(\alpha_2), \ldots$, of the obvious Skolem hull construction starting with $L(\alpha + 1)$, where $\alpha = \alpha_0 < \alpha_1, \ldots$, and each α_{i+1} is picked minimally greater than α_i, and $k \geq 0$. Note that we apparently cannot prove that this exists for all $k \geq 0$. However, if it exists for k then it exists for $k+1$, and the construction for $k+1$ extends the construction for k.

By replacement, we obtain β such that the above finite sequences that exist all end with an ordinal $< \beta$. Now we can use foundation to conclude that there are such sequences of every finite length, which can be put together into an infinite sequence with limit $\lambda \leq \beta$. Then $L(\lambda)$ is the desired partial elementary substructure of L. This suffices to establish separation since the standard integer n is arbitrary.

Finally, we verify power set. Fix $L(\alpha)$, $\alpha \geq \omega$, and suppose that new subsets of $L(\alpha)$ appear arbitrarily high up in the constructible hierarchy. Since there is a bijection from $L(\alpha)$ onto α, we see that new subsets of α appear arbitrarily high up in the constructible hierarchy. Let B be the unbounded class of all ordinals $\beta > \alpha$ such that $L(\beta + 1) \backslash L(\beta)$ has an element from $L(\alpha)$. By ω replacement, B has a closed unbounded class C of limits, none of which lie in B.

We define $A_\lambda \subseteq \lambda$, $\lambda \in C$, as the first subset of α, in the constructible hierarchy, lying in $L(\beta + 1) \backslash L(\beta)$, where β is the least element of B greater than λ. Since the A_λ are all different subsets of α, we have a counterexample to the subtlety of On. □

Lemma 60. ZFC + $V = L$ + SVWISUB *is interpretable in* $\sigma + \#$.

Proof. This follows easily from Lemma 59. □

We now wish to verify that # holds in $(L, \in_L)$. By Lemma 57, σ holds in $(L, \in_L)$. This will provide an interpretation of $\sigma + \#$ in Newcomp + Ext.

In ordinary set theoretic terms, the hypothesis for # is

H1) an assignment $A_\alpha \subseteq \alpha$, for all ordinals α, where the strict sup of every A_α is α.

The conclusion is

C1) the existence of $2 \leq \alpha < \beta$ such that $A_\alpha \subseteq A_\beta$.

Suppose the hypothesis H1) holds in $(L, \in_L)$. This gives us a virtual assignment $A_\alpha \subseteq \alpha$, for all "ordinals" α, where we have to be careful about what notion of ordinal is being used here. These will be the "ordinals" as given by pairs (x, d), where $W'(x)$

and $d \in x$. These are "measured" according to the position of d in $\leq(x)$. I.e., we are factoring by the equivalence relation $\equiv(x)$. Of course, the same "ordinal" is also "measured" by pairs (y, e), $W'(y)$, provided the position of e in $\leq(y)$ is the same as the position of d in $\leq(x)$. These positions are compared by COMP(x, y).

A "set of ordinals" is therefore given by pairs (x, u), where $W'(x)$ and $u \subseteq x$. Here we require that u is closed under the equivalence relation $\equiv(x)$. There is the obvious relationship between another such pair (y, v), representing the same "set of ordinals".

Thus H1) is represented by

H2) a virtual function F that maps the pairs (x, d), $W'(x), d \in x$, to sets $F(x, d) \subseteq x_{<d}$, where $F(x, d)$ is closed under $\equiv(x)$ and the strict sup of $F(x, d)$ is d in the sense of $\leq(x)$. The function F will produce the same "set of ordinals" at other (x', d') that represent the same "ordinal".

Assume that C1) fails in $(L, \in_L)$. I.e.,

H3) all "inclusions" among the A_α have $\alpha =$ "0" or "1", all in the sense of $(L, \in_L)$.

We need to appropriately interpret this statement using Lemma 31. The interpretation is that

H4) if $F(x, d) \subseteq F(x, e)$ then $d = \emptyset$ or $d = \{\emptyset\}$.

To complete the contradiction, we now use this data to build a giant transitive set, with a system, that goes through the roof of L.

This situation is a more exotic form of the construction used in Lemma 37 that starts with an x-system S and defines $S\#$ and S^*. Also see the paragraph after Lemma 37.

It is convenient to rearrange F so as to assign an x-system to each x with $W'(x)$. Thus we have the following, which heavily uses separation.

H5) Let G be the virtual function such that for each x with $W'(x)$, $G(x)$ is the x-system $R \subseteq x \cdot x$, where $R(a, b) \longleftrightarrow b \in F(x, a)$. We have the following crucial coherence property. Let y be such that $W'(y)$. Then either the x-system $G(x)$ is isomorphic to a proper initial segment of the y-system $G(y)$, or the y-system $G(y)$ is isomorphic to a proper initial segment of the x-system $G(x)$, or the x-system $G(x)$ is isomorphic to the y-system $G(y)$, where, in any case, the virtual isomorphism relation is COMP(x, y).

We emphasize that, even though F, G are virtual, when localized to x with $W'(x)$, they become actual in view of separation (i.e., every virtual subset of a set is a set).

We now use the construction arising out of Lemma 37. We consider the various $G(x)\#$ and their ranges $G(x)^*$. Again, these cohere using the COMP(x, y). In particular, the $G(x)^*$ cohere in the following strong sense. For any two $G(x)^*, G(y)^*$, one is an actual initial segment of the other. This is formulated using $\leq(G(x)^*), \leq(G(y)^*)$. So when we take primes ($'$), the relevant isomorphism relations are identity functions. In light of H4) or Lemma 35, we see that any proper inclusion among these $G(x)^*$ has left side $\emptyset$ or $\{\emptyset\}$.

To verify that # holds in $(L, \in_L)$, it suffices to derive a contradiction from the hypothesis H5).

Lemma 61. *Assume the hypothesis* H5). *Let E be the virtual union of the $G(x)^*$. Then E is a transitive virtual set. Any proper inclusion among elements of E has left side $\emptyset$ or $\{\emptyset\}$. E is well founded. E is nonempty. $W'(E)$. For all x with $W'(x)$,* COMP(x, E) *maps all of x.*

Proof. The transitivity of E is immediate since it is the union of transitive sets. Any proper inclusion among elements of E is a proper inclusion among elements of some $G(x)^*$ by coherence. Hence by Lemma 35, it must have left side $\emptyset$ or $\{\emptyset\}$. Hence all expansions in E meet $\{\emptyset, \{\emptyset\}\}$. Therefore E is a set. Now E is well founded because it is the union of well founded transitive sets. E is nonempty because $W'(\omega\hat{})$ and $G(\omega\hat{})^*$ is nonempty. $\leq(E)$ is a direct limit of the $\leq(G(x)^*)$, each one of which has no maximum element. Hence $\leq(E)$ has no maximum element. $W'(E)$ has now been established. For the final claim, if $W'(x)$ then $G(x)^*$ is an initial segment of E. Hence COMP(x, E). □

Lemma 62. *The last two claims of Lemma* 61*, $W'(E)$, and for all x with $W(x)$,* COMP(x, E) *maps all of x, are jointly impossible.*

Proof. By Lemma 52, we can find x with $W'(x)$, where COMP(x, E) maps a proper initial segment of x onto E. This violates trichotomy. □

Lemma 63. H5) *is false.* # *holds in* $(L, \in_L)$.

Proof. The first claim is by Lemma 62. The second claim is by the first claim and the paragraph just before Lemma 61. □

Theorem 64. ZFC + $V = L$ + SVWISUB *is interpretable in* Newcomp + Ext. ZFC + $V = L$ + SSUB *is interpretable in* Newcomp.

Proof. By Lemma 60, ZFC + $V = L$ + SVWISUB is interpretable in $\sigma + \#$. By Lemmas 57 and 63, $\sigma + \#$ holds in $(L, \in_L)$, and this has been shown in Newcomp+Ext. Hence $\sigma + \#$ is interpretable in Newcomp + Ext. This establishes the first claim. For the second claim without $V = L$, use the first claim together with Theorem 21 and Corollary 7. By standard relativization to L, we can add $V = L$. □

Theorem 65. Newcomp, Newcomp + Ext, ZFC + SSUB *are mutually interpretable.*

Proof. ZFC + SSUB is interpretable in Newcomp by Theorem 64. Newcomp +Ext is interpretable in Newcomp by Theorem 21. Newcomp is interpretable in ZFC + SSUB by Theorem 8. □

Acknowledgement. This research was partially supported by NSF Grant DMS-9970459.

References

[1] Russell, Bertrand: 1902. Letter to Frege. In: J. van Heijenoort (ed.), *From Frege to Gödel*, Cambridge, MA: Harvard University Press, 1971, 124–125.

[2] Baumgartner, James: 1975. Ineffability properties of cardinals I. In: A. Hajnal *et al.* (eds.), *Infinite and Finite Sets*, Colloquia Mathematica Societatis Janos Bolylai vol. 10, Amsterdam: North-Holland, 109–130.

[3] Baumgartner, James: 1977. Ineffability properties of cardinals II. In: R. E. Butts and J. Hintikka (eds.), *Logic, Foundations of Mathematics and Computability Theory*, Dordrecht: Reidel, 87–106.

[4] Friedman, Harvey: 2001. Subtle cardinals and linear orderings. *Annals of Pure and Applied Logic* 107: 1–34.

[5] Friedman, Harvey: 2003. Primitive independence results. *Journal of Mathematical Logic* vol. 3, No. 1: 67–83.

[6] Kanamori, Akihiro: 1994. *The Higher Infinite*. Perspectives in Mathematical Logic. Berlin: Springer.

The Ohio State University
Mathematics Building
231 West 18th Avenue
Columbus, OH 43210
USA

E-mail: friedman@math.ohio-state.edu

Completeness and Iteration in Modern Set Theory

Sy D. Friedman

Abstract. As the result of work in modern set theory, a very attractive picture of the universe of sets is starting to emerge, a picture based upon the existence of inner models satisfying large cardinal axioms. In this article I shall argue for the correctness of this picture, using the principles of *completeness* and *iteration*.

Set theory entered the modern era through the work of Gödel [8] and Cohen [2]. This work provided set-theorists with the necessary tools to analyse a large number of mathematical problems which are unsolvable using only the traditional axiom system ZFC for set theory. Through these methods, together with their subsequent generalisation into the context of large cardinals, set-theorists have had great success in determining the axiomatic strength of a wide range of ZFC-undecidable statements, not only within set theory but also within other areas of mathematics.

Through this work a very attractive picture of the universe of sets is starting to emerge, a picture based upon the existence of inner models satisfying large cardinal axioms. In this article I shall argue for the correctness of this picture, using the principles of *completeness* and *iteration*.

1. Constructibility

Gödel [8] provided an interpretation of ZFC whose structure can be thoroughly analysed. The universe L of *constructible sets* consists of all sets which appear within the hierarchy

$L_0 = \emptyset$,

$L_{\alpha+1} =$ The set of definable subsets of L_α,

$L_\lambda = \bigcup\{L_\alpha \mid \alpha < \lambda\}$ for limit λ,

$L = \bigcup\{L_\alpha \mid \alpha \in \mathrm{ORD}\}$.

This hierarchy differs from the von Neumann hierarchy of V_α's in that only *definable* subsets are considered at successor stages, as opposed to arbitrary subsets. By restricting the power set operation in this way, one brings the notion of set much closer

to that of ordinal number, and can achieve as clear an understanding of arbitrary sets as one has of ordinal numbers. As an example, Gödel showed that in L the set of reals and the set of countable ordinals have the same cardinality.

Jensen [9] went one step further, by dividing the transition from L_α to $L_{\alpha+1}$ into ω intermediate levels $L_\alpha \subseteq L^1_\alpha \subseteq L^2_\alpha \subseteq \cdots \subseteq L_{\alpha+1}$. The power of Jensen's idea, which led to his *fine structure theory for* L, is that these new successor levels L^{n+1}_α consist of sets which can be enumerated in a way analogous to that in which the Σ^0_{n+1}-definable sets of arithmetic can be recursively enumerated using an oracle for the n-th Turing jump $0^{(n)}$. What is perhaps surprising is that this "ramification" of levels can be used to establish new results concerning the structure of the constructible universe as a whole. For example, Jensen showed that the following combinatorial principle holds in L:

The $\square$ *Principle.* To every limit ordinal α that is not a regular cardinal, one can assign an unbounded subset X_α of α with ordertype less than α, such that if $\bar{\alpha}$ is a limit of elements of X_α then $X_{\bar{\alpha}} = X_\alpha \cap \bar{\alpha}$.

Every known proof that this principle holds in L makes use of some version of the fine structure theory. Indeed, Jensen's theory is so powerful that one has the impression that any question in combinatorial set theory (not involving the consistency of ZFC) can be resolved under the assumption $V = L$.

L is the least *inner model*, i.e., transitive class containing all ordinals, in which the axioms of ZFC hold. Can we construct larger inner models which admit a similar Gödel–Jensen analysis?

2. Completeness

It will be convenient to work now, not with the usual theory ZFC, but with the Gödel–Bernays theory of classes GB. This theory is no stronger than ZFC, but allows us to discuss classes which may not necessarily be definable. For an inner model M, a class *A belongs to* M iff $A \cap x$ belongs to M for every set x in M.

$V = L$ (i.e., the statement that every set is constructible) is not a theorem of GB: The forcing method allows us to consistently enlarge L to models $L[G]$ where G is a set or class that is P*-generic over* L for some L*-forcing* P, i.e., some partial ordering P that belongs to L. Thus it is consistent with GB that there are inner models larger than L.

Assume now that generic extensions of L do exist, and let us see what implications this has for the nature of the set-theoretic universe. For this purpose we introduce the notion of *CUB-completeness*.

Definition. A class of ordinals is CUB (closed and unbounded) iff it is a proper class of ordinals which contains all of its limit points. A class X of ordinals is *large* iff it contains a CUB subclass.

Largeness is not absolute: It is possible that a class X belonging to L is not large but becomes large after expanding the universe by forcing.

Definition. A class X is *potentially large* iff it is large in a generic extension of the universe.

Now we pose the following question: Can the universe be complete with respect to the largeness of classes that belong to L? That is, can the universe be *CUB-complete over L* in the sense that every class which belongs to L and is potentially large is already large? Using the fact that Jensen's $\square$ Principle holds in L, we have the following.

Theorem 1 ([5]). *There exists a sequence X_n, $n \in \omega$ of classes such that:*

1. *Each X_n belongs to L and indeed the relation "α belongs to X_n" is definable in L.*
2. *$X_n \supseteq X_{n+1}$ for each n and each X_n is potentially large.*
3. *If each X_n is large then the universe is* CUB*-complete over L.*

Thus we have the following picture: Let n be least so that X_n is not large, if such a finite n exists, and $n = \infty$ otherwise. If n is finite then n can be increased by going to a generic extension of the universe, further increased by going to a further generic extension, and so on. The only alternative is that the universe be CUB-complete over L.

Can there be many different ways of making the universe CUB-complete over L? The next result says not.

Theorem 2 ([5], [12]). *If the universe is* CUB*-complete over L then there is a smallest inner model $L^{\#}$ which is* CUB*-complete over L.*

Thus $L^{\#}$ is the "canonical" completion of L with respect to largeness of classes that belong to L.

What is $L^{\#}$? This model is *not* a generic extension of L, but rather a new kind of extension, which can be defined in terms of the concept of *rigidity*: An *embedding of* L is an elementary embedding $\pi : L \to L$ which is not the identity. We say that L is *rigid* iff there is no such embedding.

Fact (Kunen, see [12]). The universe is CUB-complete over L iff L is not rigid. If this is the case then $L^{\#}$ is the smallest inner model to which an embedding of L belongs.

A more explicit description of $L^{\#}$ is the following: Let $\pi : L \to L$ be an embedding of L and let α_π be the least ordinal such that $\pi \restriction \alpha$ is not an element of L. Then $L^{\#} = L[\pi \restriction \alpha_\pi]$ for any choice of π.

$L^{\#}$ is usually written as $L[0^{\#}]$, where $0^{\#}$ is a special set of integers [13], and the hypothesis that L is not rigid is usually written as "$0^{\#}$ exists".

The hypothesis that $0^{\#}$ exists not only completes the universe with regard to the largeness of classes that belong to L, but also solves another mystery: It can be shown that P-generics cannot exist simultaneously for all L-forcings P. How can we decide whether or not a P-generic should exist for a given P? The next result leads us to a good criterion, under the assumption that $0^{\#}$ exists.

Theorem 3 ([4]). *Assume slightly more than* GB (*precisely:* ORD *is* $\omega + \omega$*-Erdős*). *If* $0^{\#}$ *exists,* P *is an* L*-forcing which is definable in* L (*without parameters*) *and there exists a* P*-generic, then there exists a* P*-generic definable in* $L[0^{\#}]$.

Thus the inner model $L[0^{\#}]$ is *saturated* with respect to L-definable forcings. The existence of P-generics for L-definable forcings P is thereby resolved by the assumption that $0^{\#}$ exists: P has a generic iff it has one definable in $L[0^{\#}]$.

3. Iteration

The above discussion leads us to the existence of $0^{\#}$. We can go further: $0^{\#\#}$ relates to the model $L[0^{\#}]$ in the same way as $0^{\#}$ relates to L, and its existence follows from the CUB-completeness of the universe with respect to $L[0^{\#}]$. Indeed, through iteration of a suitable "# operation", we are led to models much larger than L, which satisfy strong large cardinal axioms.

We have said that the existence of $0^{\#}$ is equivalent to the non-rigidity of L, i.e., to the existence of an embedding of L. Let us use this as a basis for generalisation. Suppose that M is a non-rigid inner model and let $\pi : M \to M$ be an embedding of M. Let κ be the *critical point of* π, i.e., the least ordinal such that $\pi(\kappa) \neq \kappa$. (For technical reasons we assume that M is of the form L^A for some class of ordinals A and that π *respects* A in the sense that $\pi(A \cap \kappa) = A \cap \pi(\kappa)$, $H^M_\kappa = L^A_\kappa$ and $H^M_{\pi(\kappa)} = L^A_{\pi(\kappa)}$.) For some least ordinal α, the restriction $\pi \restriction \alpha$ is not an element of M. Normally this ordinal is κ^+ of M. We then define the # (or *extender*) *derived from* π to be the restriction $E_\pi = \pi \restriction (\kappa^+ \text{ of } M)$. A # *for* M is a # derived from some embedding $\pi : M \to M$. Thus M has a # iff M is non-rigid.

A # *iteration* is a sequence $M_0, M_1, \ldots$ of inner models where

$M_0 = L$

$M_{i+1} = M_i[E_i]$, where E_i is a # for M_i

M_λ for limit λ is the "limit" of $\langle M_i \mid i < \lambda \rangle$.

The type of model that arises through such an iteration is called an *extender model* and is of the form $L[E]$ where $E = \langle E_\alpha \mid \alpha \in \text{ORD} \rangle$ is a sequence of extenders, see [3].

How large an extender model can we produce through #-iteration? A #-iteration is *maximal* if it cannot be continued to a larger extender model. An extender model is *maximal* if it is the final model of a maximal #-iteration. Do such models exist, and if so, how large are they?

Theorem 4 ([7]). *There exists a maximal extender model, unless there is an inner model with a superstrong cardinal.*

Superstrength is a very strong large cardinal property, more than sufficient to carry out most applications of large cardinals in combinatorial and descriptive set theory. Thus if we are interested in showing that there are inner models which satisfy useful large cardinal axioms, we may without harm assume that there is a maximal extender model.

Now we turn to the question of largeness for maximal extender models. There are two ways in which an extender model M can be maximal: Either there is no # for M, i.e., M is rigid, or M is not enlarged through the addition of a # for itself. The latter possibility also leads to an inner model with a superstrong cardinal.

Theorem 5 ([7]). *Suppose that M is a non-rigid, maximal extender model. Then in M there is a superstrong cardinal.*

Thus to obtain an inner model with a superstrong cardinal, it suffices to build a maximal extender model and then argue that it is not rigid.

But first we must address a central problem in the theory of extender models: How can we ensure that our maximal extender models satisfy the Gödel and Jensen properties, GCH and $\square$? Define an extender model to be *good* iff it satisfies GCH and $\square$. The construction of *good* maximal extender models is far more difficult than the construction of arbitrary maximal extender models. However using work of Steel [14] and Schimmerling-Zeman [11] we do have the following, assuming slightly more than GB (precisely: ORD is subtle and weakly compact).

Theorem 6. *There is a* good *maximal extender model, unless there is an inner model with a Woodin cardinal.*

Thus if we can argue for the non-rigidity of good maximal extender models, we obtain an inner model with a Woodin cardinal, a property still strong enough to carry out many applications of large cardinals to combinatorial and descriptive set theory.

4. Completeness, Again

Using CUB-completeness we argued that L is non-rigid. Can this argument be generalised to good maximal extender models?

Theorem 7 ([7]). *Suppose that the universe is* CUB*-complete over a good maximal extender model. Then there is an inner model with a measurable cardinal.*

Measurable cardinals are much stronger than $0^{\#}$, but still far weaker than Woodin cardinals. To go further, we need to consider a variant of CUB-completeness.

Definition. A class of ordinals C is CUB^+ iff C is CUB and for each limit cardinal α in C, $C \cap \alpha^+$ is CUB in α^+. X is large$^+$ iff X contains a CUB^+ subclass. X is *potentially* large$^+$ iff X is large$^+$ in a generic extension of the universe.

Now we repeat what we did earlier for L, with CUB-completeness replaced by CUB^+-completeness.

Theorem 8. *Suppose that M is a good maximal extender model. Then there exists a sequence X_n, $n \in \omega$ of classes such that:*

1. *Each X_n belongs to M.*
2. *$X_n \supseteq X_{n+1}$ for each n and each X_n is potentially* large$^+$.
3. *If each X_n is* large$^+$ *then for some* CUB *class C: $(\alpha^+$ of $M) < \alpha^+$ for α in C.*

Thus if the universe is CUB^+-complete over a good maximal extender model M, it follows that α^+ of M is less than α^+ for α belonging to a CUB class. Now we apply the following refinement of Theorem 6, see [14].

Theorem 9. *Unless there is an inner model with a Woodin cardinal, there is a good maximal extender model M with the following property: For no* CUB *class C is $(\alpha^+$ of $M) < \alpha^+$ for α in C.*

Putting this all together:

Theorem 10. *Assume slightly more than* GB (*precisely:* ORD *is subtle and weakly compact*). *Suppose that the universe is* CUB^+*-complete with respect to good maximal extender models. Then there is an inner model with a Woodin cardinal.*

In the sense of Theorem 10, completeness and iteration can be used to argue in favour of the existence of inner models with large cardinals.

5. Speculations Beyond One Woodin Cardinal

Further work of Andretta, Jensen, Neeman and Steel (see [1], [10]) suggests that a good maximal extender model should exist, unless there is an inner model with many Woodin cardinals. But there have been serious obstacles to extending this work up to the level of a superstrong cardinal.

Perhaps the difficulties come from definability. The good extender models M that have been constructed until now satisfy:

$(*)$ M is a definable inner model (in the language of class theory).

Property $(*)$ is not used in our above discussion of CUB and CUB^+ completeness, nor in many applications of inner models, and may have to be sacrificed if one is to reach the level of a superstrong cardinal.

One way to approach this question is to ask: Assume that a superstrong cardinal exists. What kinds of inner models with a superstrong cardinal can one construct?

Conjecture. Suppose that there is a superstrong cardinal. Then:

(a) There is an inner model of GCH which has a superstrong cardinal.

(b) There need not be a definable such inner model.

If true, this conjecture implies that one should look not for "canonical" inner models for large cardinals, but rather a family of inner models, any of which could serve as a good approximation to the universe of all sets.

Acknowledgement. The author was supported by NSF Grant Number 9625997-DMS.

References

[1] Andretta, Alessandro, Itay Neeman, and John Steel: to appear. The domestic levels of K^c are iterable. *Israel Journal of Mathematics*.

[2] Cohen, Paul: 1963. The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences, USA* vol. 50: 1143–1148.

[3] Dodd, Anthony: 1982. *The Core Model*. London Mathematical Society Lecture Note Series, vol. 61, Cambridge University Press.

[4] Friedman, Sy: 1998. Generic saturation. *Journal of Symbolic Logic* vol. 63 (1): 158–162.

[5] Friedman, Sy: 1999. New Σ^1_3 facts. *Proceedings of the American Mathematical Society* 127: 3707–3709.

[6] Friedman, Sy: 2000. *Fine Structure and Class Forcing*. De Gruyter Series in Logic and its Applications, vol. 3. Berlin: De Gruyter.

[7] Friedman, Sy: 2001–2002. *Topics in Set Theory*. Lecture Notes, Institut für Formale Logik, Universität Wien.

[8] Gödel, Kurt: 1940. *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*. Annals of Mathematics Studies # 3. Princeton: Princeton University Press.

[9] Jensen, Ronald B.: 1972. The fine structure of the constructible hierarchy. *Annals of Mathematical Logic* vol. 4: 229–308.

[10] Neeman, Itay: 2002. Inner models in the region of a Woodin limit of Woodin cardinals. *Annals of Pure and Applied Logic* vol. 116 (1–3): 67–155.

[11] Schimmerling, Ernest and Martin Zeman: 2001. Square in core models. *Bulletin of Symbolic Logic* vol. 7 (3): 305–314.

[12] Silver, Jack: 1971. Some applications of model theory in set theory. *Annals of Mathematical Logic* vol. 3 (1): 45–110.

[13] Solovay, Robert: 1967. A nonconstructible Δ^1_3 set of integers. *Transactions of the American Mathematical Society* vol. 127: 50–75.

[14] Steel, John: 1996. *The Core Model Iterability Problem*. Lecture Notes in Logic 8. Berlin: Springer.

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

E-mail: sdf@math.mit.edu

Was sind und was sollen (neue) Axiome?

Kai Hauser

Abstract. The paper examines some philosophical issues surrounding the ongoing search for new axioms of set theory in the light of recent mathematical developments.

1. The standard axioms of set theory ZFC (Zermelo Fraenkel with the axiom of choice) are successful in several respects. They provide a formal description of the intuitive notion of a *collection* that is amenable to mathematical analysis. They also serve as a foundation of mathematics in the following sense.

- All mathematical statements can be expressed in the language of set theory with *set* and *membership* as the only primitive terms.[1]
- Every theorem of classical mathematics is provable in ZFC.

Nevertheless these axioms leave open many natural and fundamental questions. The most famous one concerns the cardinality of the continuum (the set of real numbers). Cantor's first publication on the theory of sets [4] contained a proof that the continuum is uncountable. Unable to determine its exact cardinality, Cantor subsequently conjectured that the continuum is of size the first uncountable cardinal, i.e., the minimal value permitted by the aforementioned theorem.[2] Under the name *continuum hypothesis* (CH), Cantor's conjecture soon became the outstanding problem of the subject, and it appeared as the first item on Hilbert's famous list of unsolved mathematical questions [28]. Roughly four decades later, Gödel [18] showed that ZFC, if consistent, cannot refute CH.[3] Finally, Cohen [10] established that CH is not provable in ZFC, again assuming the consistency of the latter.[4] The techniques developed by Gödel and Cohen yield many more independent propositions including propositions from other areas of mathematics not containing set theoretic vocabulary.[5] One of the

[1] Formally speaking, $\in$ is the only extra-logical symbol of the language of set theory. But from the epistemological point of view, 'set' and 'membership' are both autonomous owing to the difference in the underlying cognitive acts.

[2] The conjecture appears in print for the first time in [5]: 132.

[3] Gödel proved that if ZF (Zermelo Fraenkel *without* the Axiom of Choice) is consistent, then so is ZFC together with a global form of CH.

[4] The assumption that ZF is consistent suffices in both cases by virtue of the first half of footnote 3.

[5] Some examples are Suslin's hypothesis [41], the Kaplanski conjecture [11], the Whitehead problem [51] and the S- and L-space problems in general topology [56].

main motivations for the introduction of new axioms has been the need to settle these open problems. This could mean in particular that the proposition in question or its negation is logically derivable from a new axiom. But more subtle interpretations of 'settle' are conceivable, as we shall see below. In either case, we must ask why those statements that are proposed as new axioms really are axioms, and moreover, what we mean by 'axiom' in the first place.

The paper aims to shed some light on these questions. It is organized as follows: Sections 1 through 6 outline the mathematical background and place it into a philosophical context taking the philosophical views of Gödel as a point of departure.[6] This forms the basis of the philosophical argument contained in Sections 7 through 10. Although it falls short of answering the above questions, that argument indicates the direction where the answers may be found.

2. An influential proposal to search for new axioms was advanced in [20] along with the suggestion that the role of the continuum problem in set theory is "that it will finally lead to the discovery of new axioms which will make it possible to disprove Cantor's conjecture" (ibid., 186). Anticipating independence from ZFC, Gödel defended the existence of a determinate truth value for CH on the grounds that the concepts of classical set theory describe some well-determined reality.

> For in this reality Cantor's conjecture must be either true or false, and its undecidability from the axioms as known today can only mean that these axioms do not contain a complete description of this reality. ([20]: 181)

This immediately raises two questions: What exactly is this reality described by the axioms of set theory, and how do we recognize *axioms* yielding a more detailed description? Gödel subscribes to a distinctive brand of platonism modeling the existence of sets as mathematical *objects* as well as our epistemic access to them in direct analogy with physical objects and sense perception. At the same time he maintains that the objects of transfinite set theory "clearly do not belong to the physical world, and even their indirect connection with physical experience is very loose" ([23]: 267). That makes it doubtful whether the causal theory of sense perception can provide a model for the epistemology of set theory. Nevertheless, in the next paragraph Gödel insists that mathematical intuition is a kind of perception of abstract objects enabling us to formulate mathematical axioms. But he does not explain how these objects determine our intuitions.

One way of circumventing these difficulties is to suspend judgement about ontology, and to focus directly on epistemological matters. To begin with, it is undeniable that we are in possession of relatively stable intuitions about the concept of set as articulated by the ZFC axioms. With the exercise of reason new intuitions emerge and

[6] My reasons for choosing this somewhat biased approach are twofold. First, in his time Gödel had the most deeply thought-out point of view on the philosophical foundations of set theory. Second, the mathematical developments which are most relevant to the paper are to a large extent an outgrowth of Gödel's work.

in turn give rise to additional candidates for axiomhood. In spite of the psychologist flavor, the key point here is that the whole process is anything but arbitrary. On the one hand various formal constraints need to be obeyed (e.g., logical consistency). On the other hand there is a kind of 'informal rigor' involved that is much harder to pin down and is implicit in locutions like 'principle *A* is true by virtue of the intended meaning of the concept of set'. An indication of its importance is that the standards governing the acceptance of new *axioms* are *de facto* rather strict. Given our ability to successively refine mathematical intuitions in a 'rule-like' fashion, we may treat statements which are undecidable in ZFC as determinately true or false without reference to the state of affairs in a realm of mathematical objects.[7] In effect this line of thought is adopted in the supplement to the second edition of Gödel's article of the continuum problem.

> However, the question of the objective existence of *objects* of mathematical intuition (which, incidentally, is an exact replica of the question of the objective existence of the outer world) is not decisive for the problem under discussion here. The mere psychological fact of the existence of an intuition sufficiently clear to produce the axioms of set theory and an open series of extensions of them suffices to give meaning to the question of the truth or falsity of propositions like Cantor's continuum hypothesis. ([23]: 268, emphasis added)

Actually Gödel held that the open-endedness of attempts to axiomatize the concept of set is itself intrinsic to this concept.

> For first of all the axioms of set theory by no means form a system closed in itself, but, quite on the contrary, the very concept of set [...] on which they are based suggests their extension by new axioms which assert the existence of still further iterations of the operation "set of". ([20]: 181, [23]: 260)[8]

The concept Gödel had in mind is the *iterative concept* of set "according to which a set is anything obtainable from the integers (or some other well-defined objects) by iterated application of the operation 'set of', and not something obtained by dividing the totality of all existing things into two categories."[9] In formal terms the iterative concept leads to a stratification of the set theoretic universe into a cumulative hierarchy of stages V_α indexed by ordinal numbers. It begins with $V_0 = \emptyset$. In moving from one stage to the next all subsets of the previous stage are collected, i.e., $V_{\alpha+1}$ equals the power set of V_α. At limit stages λ, we take unions, i.e., $V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$. The axiom of foundation guarantees that any set appears in one of these levels. In other words the class of all sets V (formally the class term $\{x : x = x\}$) is the union of the V_α taken over all ordinals. One may thus visualize the universe as having the shape of a

[7] For an elaboration of this viewpoint see [25].

[8] See also [21]: 306f.

[9] Gödel expressively allows transfinite iterations of the operation 'set of'. He also remarks that "as opposed to the concept of set in general (if considered as primitive) we have a clear notion of this operation" ([20]: footnotes 12 and 13, 180).

funnel with bottom end $V_0 = \emptyset$ and with the ordinals marking its height as illustrated in Figure 1.

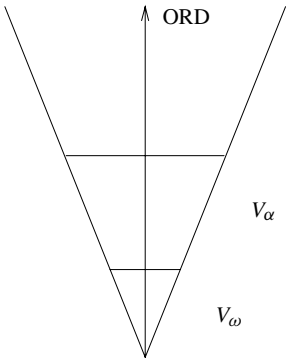


Figure 1. The cumulative hierarchy

How does this lead to new axioms asserting the existence of further iterations of the operation 'set of'? One possible answer is that implicit in the concept of set is the fact that any operation on sets propagates its own iteration. For example, an operation F acting on individual sets immediately gives rise to a new operation $\tilde{F}$ associating to a collection of sets the collection of their images under F. The universe V is 'closed' under the passage from F to $\tilde{F}$. (For definable F this is the replacement axiom.) However, it would be unreasonable to expect that V is the only closure point, for this would run counter to the intuition behind Cantor's *Absolute*— the doctrine that the set theoretic universe comprehends all possibilities.[10]

Roughly speaking, if something happened 'for the first time' in V we should not have stopped there, and the process of generating sets should have kept going. It follows there are many levels allowing to pass from a given F to the associated $\tilde{F}$. The first such level is V_ω which consists of all sets built up in finitely many stages from the empty set. Uncountable indices of such levels are called *inaccessible cardinals* indicating that they are unreachable from below by certain operations on smaller cardinal numbers.[11] The postulation of such indices is a *large cardinal axiom* by virtue of their transcendence over smaller cardinals and in the sense that their existence is unprovable in ZFC.

[10] In metaphysical terms, Cantor [6] (cf. [9]: 175, 205) maintained that "the true Infinite or Absolute does not permit any determination", a view he attributes also to Locke, Descartes, Spinoza and Leibniz. In addition he mentions Nicolaus Cusanus who had actually put forward this view in *De Docta Ignorantia*. From the Absolute Cantor distinguishes another form of the actually-given infinite calling it the *the Transfinite*. Manifestations of the latter include cardinal numbers and order types that can be grasped in mathematical thought and are capable of augmentation. The Absolute, by contrast, is "realized in highest perfection, in wholly independent, other-worldly being, *in Deo*" ([7]: 378).

[11] The two operations are cardinal exponentiation and taking suprema over families of cardinals.

After further extrapolation one arrives at *reflection principles*. Here the idea is that any property holding in V must already be true in suitable initial segments V_α. Levy [42] and Bernays [2] showed that the standard axioms and some of their generalizations can be recast as reflection principles. Suitably formulated reflection principles also prove the existence of Mahlo cardinals and illuminate their schematic arrangement into a hierarchy studied originally in [44] and [45] via thinning operations on inaccessible cardinals. As long as the reflecting formulas are only allowed to contain set parameters, the corresponding reflection principles must themselves reflect to initial segments of V. When second order parameters are allowed this yields *indescribable cardinals*[12] which again fall into a systematic hierarchy, this time with a qualitatively new form of transcendence in terms of definable filters. The mathematical results obtained in this area so far vindicate Gödel's claim that

> these axioms show clearly, not only that the axiomatic system of set theory as known today [i.e., ZFC] is incomplete, but also that it can be supplemented without arbitrariness by new axioms which are only the natural continuation of the series of those set up so far. ([20]: 182)

3. The large cardinals studied in modern set theory—*measurable, strong, Woodin* and (*super*)*compact cardinals* and their relatives—are of much larger order than those mentioned up to this point. They are defined in terms of *elementary embeddings*, truth-preserving transformations of the universe of all sets. Attempts to justify the existence of such cardinals via reflection principles have so far not been entirely successful.[13] Indeed a plausible intuition concerning elementary embeddings eventually led to the postulation of a principle contradicting the axiom of choice [40]. This principle can be viewed as the limit of a series of large cardinal axioms beginning with measurability and continuing past supercompactness.[14]

It has also been pointed out that the larger large cardinal axioms are supported by strong arguments of analogy. Typically, such arguments single out a combinatorial property of the first infinite cardinal ω and posit the existence of uncountable cardinals having an analogous property. This is usually justified by appealing to the uniform generation of sets in the cumulative hierarchy on the grounds that this uniformity should rule out the existence of 'singularities'. It is fair to say that this line of reasoning is generally regarded as carrying less credence than justifications via reflection principles.[15] For instance measurable cardinals in their traditional definition as un-

[12] Given a collection of formulas Γ, a cardinal κ is said to be Γ-*indescribable* if for any formula $\varphi(X)$ in Γ and any $A \subset V_\kappa$: if $\varphi(X)$ holds in the structure $\langle V_\kappa, \in, A\rangle$, then there is a $\lambda < \kappa$ such that $\varphi(X)$ already holds in the structure $\langle V_\kappa, \in, A \cap V_\lambda\rangle$.

[13] Arguably this series of axioms begins with cardinals derived from reflection principles as indescribable cardinals possess a reformulation in terms of elementary embeddings [24].

[14] Its original motivation, however, came from investigations of the concept of set within broad notions of class or property. See [48], [49] and also [38]: Sec. 23.

[15] Gödel, for example, held that "every axiom of infinity should be derived from the (extremely plausible) principle that V is indefinable, where definability is to be taken in [a] more and more generalized and idealized sense" ([57]: 8.7.16).

countable cardinals κ carrying a non-principal κ-complete ultrafilter may be construed as analogues of ω in virtue of the existence of non-principal ultrafilters on ω. (A filter is by definition ω-complete.) However, in this analogy there is no counterpart at ω of the crucial consequence of the existence of a non-principal, κ-complete ultrafilter on κ, namely the associated elementary embedding into a subuniverse of V. If (strongly) compact cardinals ([38]: §4) are conceived as strong analogues of ω by means of the compactness theorem for first order logic, their connection with intuitions about the concept of set in its usual meaning becomes rather tenuous.

4. Besides the intrinsic evidence provided through mathematical intuition, Gödel mentions "another (though only probable) criterion of the truth of mathematical axioms, namely their fruitfulness" in consequences ([23]: 269). Gödel was particularly interested in consequences holding in concrete and elementary domains where the meaningfulness and unambiguity of the concepts involved could hardly be doubted. To the extent that those consequences are demonstrably not obtainable without the axiom in question, this would in turn lend support to the view that the axiom is meaningful and unambiguous as well.[16] With respect to the natural numbers, this hope has not been fully realized yet, the main reason being that no example is known of a previously studied number-theoretic question whose solution requires hypotheses beyond ZFC.[17] On the other hand, with respect to second-order number theory, Gödel's view of the role of strong axioms of infinity has been vindicated. Second-order number theory is concerned with definable sets of real numbers and their characteristic properties such as Lebesgue measurability. According to their logical complexity these sets are arranged into a canonical hierarchy known as the *projective hierarchy*. ZFC yields a structure theory for the first two levels of the projective hierarchy, but gives virtually no information about its further levels. Some partial results about higher level projective sets were obtained in the 1960s from large cardinal axioms. Substantial progress came with a different class of axioms which stipulate the existence of winning strategies in infinite, two-person games of perfect information played on the integers.[18] The hypothesis that all games with projective pay-off set are determined (*projective determinacy, PD*) leads to a canonical extension of the theory of the first two levels provided by ZFC to the entire projective hierarchy which is 'complete' in the sense that all 'natural' questions become decidable.[19] Moreover, the exploration of the projective sets under PD over the last fifteen years has also resulted in a re-

[16] See [57]: 7.1.1–7.1.5, 7.4.1.

[17] There are, however, interesting examples of combinatorial statements which are provably equivalent to the (1-)consistency of large cardinal axioms. See [14], [15], and [16]. Still practitioners of number theory are reluctant to accept them as genuinely number theoretic questions on an equal footing with say the Goldbach Conjecture or problems about the distribution of prime numbers.

[18] To each $A \subset [0, 1]$ associate an infinite, two person game G_A. Players I and II alternatingly choose $n_i \in \{0, 1\}$, with I beginning with n_0. Player I wins if $\sum_{i=0}^{\infty} n_i \, 2^{-i-1} \in A$. Otherwise II wins. The set A, or rather the game G_A is *determined* if one of the players has a winning strategy, i.e., a 'rule' telling him which numbers to play with the property that he will come out winning if he follows this rule no matter what numbers the other player has chosen.

markable interaction of set theory with other areas of mathematics. On the one hand the notions and dichotomies occurring in second order number theory turned out to have meaning in disciplines ranging from harmonic analysis, Banach space theory and topological dynamics to control theory and mathematical economics. Conversely sophisticated techniques from other branches of mathematics have been applied in the solution of purely set theoretic problems and furnished new insights into the relationship of the projective sets with various subdisciplines of mathematical logic.[20] The success achieved with projective determinacy brings to mind an often quoted passage from Gödel's 1947 article on the continuum problem.

> There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems [...] that, no matter whether or not they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory. ([20]: 182n)

5. Nevertheless acceptance of PD as an *axiom* was hampered by its lack of intrinsic evidence. That difficulty was overcome when it was shown that PD is implied by large cardinal axioms. More precisely, PD follows from the existence of infinitely many Woodin cardinals [46]. Further work led to the realization that determinacy axioms and large cardinal axioms are in effect one class of axioms and not two, cf. [38] and [60]. This is highly remarkable from the epistemological point of view, for it is anything but obvious that *global* principles motivated by *a priori* intuitions about the length of the cumulative hierarchy are intimately related to *local* principles which assert something about the width of a specific level near the bottom of this hierarchy and are supported by extrinsic evidence.[21]

A key ingredient in these investigations was the possibility to build canonical models for large cardinal axioms which are minimal and whose structure can be analyzed in detail. The earliest example of a canonical model of set theory is the *constructible universe* L devised by Gödel in his proof of the consistency of the (generalized) continuum hypothesis and the axiom of choice.[22] This model is generated in stages similar to V with the crucial difference that in the successor step only the subsets of the previous stage that are definable in this structure make it into the next stage. Over the years it has become clear that ZFC delivers a 'complete' theory for L in the sense that all 'natural' ZFC questions are decidable within ZFC $+ V = L$.[23] The uniform

[19] By the incompleteness theorems there is no hope to decide 'unnatural' propositions, i.e., propositions encoding metamathematical pathologies such as Gödel sentences and consistency statements.

[20] Two of these subdisciplines are recursion theory (in connection with the global structure of Turing degrees) and model theory (topological Vaught Conjecture). For an overview and further references, see [39].

[21] The formulation of larger large cardinal axioms in terms of elementary embeddings involves cofinal segments of the universe. By contrast winning strategies in integer games can be coded as real numbers and thus belong to $V_{\omega+1}$.

[22] Working in ZF, Gödel showed that all axioms of ZFC and the (generalized) continuum hypothesis hold in L [18].

generation of the members of L which underlies this phenomenon can be generalized to construct canonical models that are big enough to accommodate various large cardinals not occurring in L. Although the construction of these models presupposes the existence of large cardinals, this still comes close to a consistency proof. For it is expected that a hidden inconsistency in the large cardinal axiom would reveal itself in the detailed structure theory of its associated model.[24]

In addition to underlining the coherence of large cardinal axioms, canonical models play a central role in determining the logical strength of these axioms. Here strength is measured in terms of relative consistency, i.e., axiom A_1 is stronger than axiom A_2 if it is provable in arithmetic that the consistency of $\mathrm{ZFC} + A_1$ implies the consistency of $\mathrm{ZFC} + A_2$. Modulo equiconsistency this defines a partial order on the class of set theoretic principles. It is an empirical fact that for 'natural' large cardinal axioms, the partial order so defined is a total order.[25] Moreover, propositions with consistency strength higher than ZFC (to the extent they have been analyzed) turned out to be equiconsistent to some large cardinal axiom—*including propositions from other branches of mathematics not containing any set theoretic vocabulary.*[26] In view of the ostensibly disparate ways in which ZFC may be extended, the arrangement of the resulting theories into a linear hierarchy is surprising. The fact that the central markers along this hierarchy are the large cardinal axioms, indicates that the latter provide the natural superstructure for the standard axioms.

This also affords an improved understanding of the claim that PD is the 'correct' axiom for second order number theory. Not only does PD yield a complete structure theory for the projective sets, it is also implied by and in fact equivalent to many combinatorial principles which on the surface have nothing to do with determinacy. Most importantly, however, second order number theory as given by PD is shared by *all* sufficiently strong extensions of ZFC. The reason for this is the curious phenomenon that propositions implying the consistency of PD actually imply PD.[27]

6. The success achieved in second order number theory with PD raises the hope that some new axiom might settle the continuum problem, a statement of third order number theory. However, here matters are more delicate because of the metamathematical

[23] The assertion that every set belongs to L is abbreviated by $V = L$. There is of course no formal definition of 'natural statement'. What is meant by this is a statement expressing an 'idea' (for example a combinatorial property) in contrast with encodings of metamathematical pathologies such as Gödel sentences or consistency statements, cf. footnote 19. The completeness of $\mathrm{ZFC} + V = L$ in this sense is an empirical fact.

[24] Thus the rationale of the inner model program resembles the idea behind Gentzen's consistency proof for Peano Arithmetic.

[25] Strictly speaking this has not yet been verified in all cases of interest, but there is no reason to doubt that this will be achieved in the long run.

[26] For example the Lebesgue measurability of all sets of real numbers (in ZF) is equiconsistent with the existence of an inaccessible cardinal (in ZFC) by Solovay [53] and Shelah [52].

[27] This hinges on the closure properties of the PD models supplied by sufficiently strong axioms of infinity. In somewhat analogous fashion, large cardinals yield ZFC models whose membership relation is the standard $\in$ relation.

limitations imposed by forcing. For example, soon after Cohen's discovery of the forcing method it became clear that (subject to modest restrictions) *no* large cardinal axiom can decide CH [43].[28] Nevertheless, new developments indicate that large cardinals are relevant albeit in subtle and unexpected ways. Perhaps the most intriguing of these is the work done recently by Woodin [58].[29] It rests on an abstraction from the metamathematical situation in second order number theory under PD to the 'next structure' after second order number theory, i.e., the subsets of ω_1 definable in $P(\omega_1)$. Assuming large cardinals, Woodin showed that if there is an axiom with a similar effect on the theory of that structure as the one exerted by PD on second order number theory, then CH must fail. Moreover, he presented an arguably plausible principle having that desired effect and constructed a canonical model where it holds. In other words, he proposed an axiom deciding CH together with an argument that there can be *no* analogous axiom leading to the opposite outcome for CH. Such an asymmetry criterion is what several principles proposed earlier had been lacking. It is worth noting that a metamathematical asymmetry is also mentioned by Gödel [23] to argue for the existence of inaccessible cardinals.

Another novelty in Woodin's approach is the switch from the standard logical consequence relation to a new logic – Ω-logic – made necessary by the metamathematical limitations imposed by forcing. The aforementioned abstraction from the metamathematics of second order number theory is conceived in terms of Ω-logic. In addition, Woodin has explored connections of Ω-logic with the large cardinal hierarchy. A reformulation of a plausible closure condition for large cardinal axioms in Ω-logic yields an 'explanation' for the linearity of a coarse version of the large cardinal hierarchy. It is of course premature to tell whether these really are explanations. Moreover, the utility of that reformulation hinges on a sweeping conjecture to the effect that Ω-logic is the strongest 'reasonable' logic which is immune to forcing. Further study will be needed to know the full effect of the theory built around Ω-logic. Similarly, it is unclear whether a *solution* to the continuum problem will emerge from Woodin's definability analysis with its metamathematical accompaniments.[30] In the end a decision about CH may not even be based on a formal proof (or refutation) from a new axiom. Conceivably the discovery of some new axiom(s) might assign a privileged role to CH (or its negation) in what is then viewed as the 'correct' theory for $P(\omega_1)$. One possible criterion for correctness might be generic invariance of the theory (its stability under forcing) in the presence of suitable large cardinals, cf. [54]: Sec. 2.8.

7. The preceding discussions provide partial answers to the two questions in the title of this article in so far as they identify some candidates that have been proposed as new axioms, and by describing their purpose primarily as a means to settle—in a manner

[28] This was noticed independently by Cohen.

[29] The two-part article [59] contains an accessible summary.

[30] Interestingly, Gödel ([20]: 182) conjectured that a more profound understanding not only of mathematical but also of logical concepts is necessary for settling CH.

that is deemed appropriate—problems left open by the ZFC axioms. But until now very little has been said about the meaning of the term 'axiom' itself.

In ordinary language the word 'axiom' is used for self-evident propositions that are accepted without proof. When it comes to mathematics, however, the criterion of self-evidence—understood as applying to propositions that are evident or known independently of all other propositions and evidence—is at best an ideal. For example, already the self-evidence of some of the standard axioms may be challenged as can be seen from the controversy surrounding the axiom of choice.[31] If we move along the large cardinal hierarchy, justifications in terms of self-evidence become more and more tenuous. Determinacy axioms lack self-evidence altogether.

These different conceptions of 'axiom' can in fact be traced back to Euclid. Among the basic principles of the *Elements* which are not definitions he distinguishes between postulates and "common notions". Only the latter assume something "which is immediately obvious and poses no difficulties to trained thought" whereas the former are not evident in themselves and are adopted mainly because they are indispensable for the theory one is aiming for.[32] When Hilbert revived Euclid's ideas at the beginning of the twentieth century he declared this indispensability as the characteristic mark of axiomhood. In [29] for example, intrinsic plausibility is not even mentioned.[33] Hilbert viewed the primary role of axioms as lending orientation and order to mathematical theorizing. Like Euclid, he does not explain what axioms are, describing instead the *axiomatic method*. Its idea is to study a mathematical theory by singling out a small mumbler of principles that cannot be reduced to more elementary ones and from which all others are logically derivable. In this manner the theory becomes organized and can be surveyed as a whole.[34] To this end axioms in the first place need to be consistent, and each axiom should be independent from the other axioms.

Hilbert's account is inadequate in several respects. For one thing, he ignores the principle that an axiom ought to have a higher degree of evidence than the theorems it proves. Another more subtle inadequacy is brought to light by projective determinacy: how could one possibly remain convinced of its consistency without believing it is *true*? Indeed this was the motivation for the program to derive determinacy from intrinsically plausible large cardinal axioms despite the overwhelming amount of extrinsic evidence supporting PD (cf. Sections 4 and 5). The higher epistemic weight of intrinsic plausibility appears to be a consequence of the fact that—in contrast with

[31] See [13] and the various views recorded in [47]. The intuitive origin of the axiom of choice is analyzed in [27].

[32] See Proclus' commentary on the first book of the *Elements*. The statement "If two things are each equal to a third thing, they must be equal to one another" is listed under the common notions whereas (an equivalent of) the parallel axiom is included among the postulates.

[33] One reason for this was no doubt Hilbert's insistence on a rather weak conception of mathematical intuition that traces back to Kant and limits the meaningful part of mathematics to finitistic reasoning, cf. [30]: 170f.

[34] Gödel's formulation "shedding so much light upon a whole field" in the passage quoted at the end of Section 4 can be understood this way. In Hilbert's view the axiomatization of theories is the basis of scientific thought and guarantees the integral unity of knowledge, see [29]: 156.

extrinsic evidence—there is at least one way in which it is directly linked with truth, namely when an axiom is 'implied' by the concept(s) it aims to capture. This comes close to the notion of analyticity described in [19]: 139, which applies when a proposition holds "owing to the meaning of the concepts occurring in it", where this meaning may perhaps be undefinable (i.e., irreducible to anything more fundamental).[35] A paradigmatic example is the axiom of foundation which is 'implied' by the iterative concept of set (cf. Section 2). Those who subscribe to this concept of set could perfectly well admit the existence of collections containing themselves as elements. Such objects would, however, not fall under the concept of set in this intended meaning.[36] By contrast, if V is to encompass all possible acts of collection and their iterations, then $V = L$ must look suspicious. One manifestation of this is that only a short initial segment of the large cardinal hierarchy is 'realized' in L. Moreover, the large cardinals not found in L furnish a combinatorial analysis of the transcendency of V over L, see [38]: ch. 2, §9. Many set theorists construe this as evidence that $V = L$ is *false*.

8. Given that the intrinsic plausibility of an axiom is inextricably bound to the meaning of the primitive terms it contains, the next question is how that meaning should be construed. Even without endorsing the thesis that meaning is determined by usage, a natural move would be to consult mathematical practice. For the manner in which debates on fundamental issues are conducted should allow conclusions about the meanings of the concepts in question. What we find in virtually all situations is that the opposing parties ultimately defend their position by invoking extra-mathematical elements in addition to mathematical facts. For example, opponents of the 'axiom' of constructibility argue that it runs counter to 'the goals of set theory' or that it violates 'the *intended* meaning of the concept of set'. What these opaque phrases suggest is that the meanings of mathematical concepts reside in the relationship in which the mind stands to the objective reality of mathematics. Moreover, if the mind is assigned a constitutive role with respect to mathematics as a whole, that relationship lies beyond the explanatory reach of mathematical facts. Similarly, arguments why extrinsic evidence really is evidence for the correctness of a given hypothesis in the last analysis say something about how the mathematician adopting this hypothesis is intentionally related to the resulting theory. Evidence, by its very nature, is a phenomenon 'occurring between' the mathematician and mathematics, and it would therefore be

[35] Gödel distinguishes from this another notion of analytic which he calls *tautological*. It is understood "in the purely formal sense that the terms occurring can be defined (either explicitly or by rules of eliminating them from sentences containing them) in such a way that the axioms and theorems become special cases of the law of identity and disprovable propositions become negations of that law" ([19]: 138n). The distinction between analytic and tautological is upheld in [21]: 321, where the comprehension scheme in second order number theory is mentioned as an axiom that is implied by the concept 'set of integers'. Gödel also points out that analytic statements need not be decidable given that our knowledge of the world of concepts is incomplete.

[36] Mathematically speaking, this is not a restriction since no interesting structures outside the wellfounded sets have been discovered.

countersensical to exclude his subjective perspective from the investigation of the role that evidence plays in the selection of axioms.[37] Admittedly, this contrasts with the concerns of 'ordinary science' where some standard of objectivity is presupposed and subjectivity is regarded as a disturbing element. There the declared goal is to abstract from the thinking subject by arranging experiments and deductions in such a way as to minimize any reference to the observer. Here we seek to understand how the subjective aspects of thought govern the rationale for selecting first principles with a claim to objective validity. In the remainder of this paper I want to outline some ideas from Husserl's phenomenology that will be helpful in this endeavor, and I will sketch their connection with the two questions asked in the title.[38]

9. Husserl, who had originally been trained as a mathematician, developed his philosophy out of a preoccupation with the epistemological foundations of mathematics. His *Logische Untersuchungen* embark on a phenomenological clarification of logic, i.e., an analysis of the essential structure of the cognitive acts by which we grasp the ideas, concepts and laws of logic. More generally, Husserl was concerned with the fundamental question of epistemology

> … how it should be *understood* that the "in itself" of objectivity becomes a "presentation" that it is "apprehended" in cognition, thus in the end nevertheless becoming subjective again; what it means that the object exists "in itself" and is "given" in cognition; how the ideality of the universal as a concept or law can enter into the flow of real psychological experiences with a claim to knowledge by the thinking subject … ([32]: 8)[39]

This problem continued to occupy a central role in the development of Husserl's thought which in its later stages is characterized by a distinctive combination of idealism and platonism. Following Kant, Husserl relies on the synthetic activity of the mind in examining how objective knowledge is possible given that experience begins with immanent, sensory contents of subjective experience. His starting point is the observation that all acts of consciousness, i.e., actual episodes of seeing, believing, hoping etc., are distinguished from other mental phenomena by a certain kind of directedness or *intentionality*: to be conscious always means to be conscious of something. Traditionally, intentionality had been conceived as a relation into which the subject enters

[37] Intuitionism is an attempt to incorporate the subjective perspective into the foundations of mathematics, but it does so by imposing unwarranted restrictions on what it regards as acceptable forms of mathematical reasoning.

[38] That there is such a connection is suggested by Gödel in a draft for an undelivered lecture [22]. The influence of Husserl's phenomenology on Gödel's philosophical program and his views on the foundations of set theory are examined in [26].

[39] … wie es denn zu *verstehen* sei, dass das "an sich" der Objektivität zur "Vorstellung", ja in der Erkenntnis zur "Erfassung" komme, also am Ende doch wieder subjektiv werde; was das heisst, der Gegenstand sei "an sich" und in der Erkenntnis "gegeben"; wie die Idealität des Allgemeinen als Begriff oder Gesetz in den Fluss der realen psychischen Erlebnisse eingehen und zum Erkenntnisbesitz des Denkenden werden kann …

with an object through a conscious act. This picture runs into problems in situations where the intended object does not exist, or an intended state of affairs does not obtain, but the corresponding act nevertheless exhibits intentionality.[40] Husserl's important insight was that the *actual* existence of an object is not decisive for the directedness of a given act. What matters for directedness is that our consciousness has a certain structure when we perform the act. Husserl calls this structure *noema*, cf. [33]: part 3, ch. 3. It is an abstract entity, distinct from both the act and its object, that can be viewed as a generalization of the concept of 'meaning' from the sphere of linguistic acts to all acts of consciousness.[41]

The function of the noema is not only to direct consciousness to objects that are already there, but also to interconnect the multifarious contents of consciousness in such a way that we have an experience *as of* an object.

> But in a wider sense an object 'constitutes' itself—'whether it is actual or not'—in certain concatenations of consciousness bearing in themselves a recognizable unity in so far as they carry with themselves by virtue of their essence the consciousness of an identical X. ([33]: §135, 281)[42]

Constitution is the main idealist ingredient in Husserl's philosophy, however, it must not be misunderstood as 'creation'. What gets constituted in consciousness is sense rather than objects, for example the sense of 'object'. By examining certain components of this sense, which Husserl called a *Leitfaden* for constitutional analyses,[43] we become acquainted with the platonist streak in his philosophy.

From the viewpoint of phenomenology objects[44] are understood primarily as invariants or identities in the constantly changing stream of consciousness. This means that awareness of an object comes about when distinct appearances are co-instantiated as appearances *of* something that remains identical throughout those appearances ([33]: §41; [34]: §58).[45] In this manner not only physical things, but also their abstract features such as shape and color become objects to which consciousness is directed. Likewise mathematical items and other idealities assume their sense of objective existence once they emerge as identities in manifolds. Two familiar examples are the continuum with its various axiomatic characterizations in topology, algebra and analysis, or the concept of a group as the common structure underlying diverse manifestations

[40] Examples are acts directed at contradictory concepts such as a round square, the set of all sets or an elementary embedding of the universe into itself under AC.

[41] This interpretation of the noema and the ensuing reading of Husserl are due to Føllesdal. Although not uncontested, it is supported by systematic and textual considerations [12].

[42] Im *weiteren* Sinne aber "konstituiert" sich ein Gegenstand – "ob er wirklicher ist oder nicht" – in gewissen Bewusstseinszusammenhängen, die in sich eine einsehbare Einheit tragen, sofern sie wesensmässig das Bewusstsein eines identischen X mit sich führen.

[43] See [36]: §§20–22, [33]: §§149, 150.

[44] In what follows the term 'object' is used in the broad sense encompassing not only physical things but also concepts, multiplicities, predicates and state of affairs etc., cf. [32]: part 2, §45, 143.

[45] Invariance and identity across different experiences also play a central role in [50]: ch. XVI and in [55]: p. 449 where they are related to Kant's ideas about empirical objects.

of symmetry. Among the large cardinal axioms, the existence of weakly compact cardinals is justified by a multitude of equivalent definitions ranging from infinitary combinatorics to compactness properties of infinitary languages, reflection principles, model theoretic ideas and algebraic structures. Woodin cardinals—whose connection with reflection principles is more tenuous than that of weakly compact cardinals—are generally regarded as definite entities because they arise in at least three seemingly unrelated contexts (stationary tower forcing, iteration trees and the extender algebra). In each of those situations Woodin cardinals encapsulate a combinatorial property of precisely the right kind to carry out the required argument, cf. [60]. The 'object-sense' of this large cardinal concept is then the noematic correlate of the act in which we apprehend what remains invariant across these different experiences: "the pure X in abstraction of all predicates" ([33]: §131).

Attesting to the platonist streak in Husserl's philosophy is his insistence that perception is inherently perspectival and incomplete. In every perceptual situation only some of the determinations of the object, for example the front side or its length, are present. Shape, color and other abstract features of the object are never fully given in a single experience. Depending on the point of view of the perceiver, they are revealed through a manifold of perspective variations (*Abschattungen*) with sufficient affinities to be apprehended as perspective variations of one and the same objective entity ([33]: §41, 74f). The aspectival nature of objects of perception is also in evidence in the manner in which mathematical entities and other idealities are experienced, making it appropriate to speak of objects and perception in this context as well. Measurable cardinals, for example, were first conceived as an affirmative answer to the abstract measure problem, cf. [38]: §2. They gained their recognition as objectively existent by virtue of the various aspects under which they can be analyzed (infinitary combinatorics, the ultrapower construction, transcendency over smaller cardinals, their effect on the constructible sets, the connection with descriptive set theory, new forcing techniques and the fine structure of the associated inner model), and the fact that measurability remains thematic throughout the numerous results obtained under each particular aspect ([37]; [38]: ch. 4).

When visual experiences are examined phenomenologically, it is found that they are not self-contained units of experience. For example, standing in front of a house I cannot see its back. Nevertheless, what I perceive is the front of something having a back which would be visible if I changed my position relative to the house. In general, it belongs to the essence of any intentional experience that we always anticipate or co-intend other determinations of the intentional object that are not in the current focus of attention. These adumbrations and tacit allusions must be *intrinsically* related across different acts as otherwise it remains unintelligible how any one of them can confirm, clarify or rule out any other. The whole system of predelineated potentialities that surrounds every act is what Husserl calls its *horizon*. It is indispensable for explaining how *within* the stream of consciousness the awareness of an enduring object is enabled ([36]: §§19, 20).

The ideas presented so far suggest how an act of consciousness can possess the intrinsic property of object-directedness irrespective of the 'actual' existence of its object. What is missing is an explanation of how this leads to *knowledge* about that object. To that end Husserl introduces the notion of *fulfillment*.[46] It describes the special mode of consciousness in which we comprehend that two qualitatively different acts fit together in a certain way. One of these acts is *signitive* and merely points to an object. For example I can name a particular number by a compound arithmetical expression like $17^2 + 29^3$ without realizing that it is the number 24 678. The other act is called *intuitive* because in its performance consciousness is explicitly presented with an object. Perceptual acts are intuitive in this sense because their objects are experienced directly and not as representations of some reality that is only mediately accessible ([33]: §43). Husserl states as a fundamental phenomenological fact that a signitive and an intuitive act may enter into a peculiar union.

> We experience how in intuition *the same* objectuality is intuitively realized which was "merely meant" in the symbolic act, and that it becomes intuitable exactly as something determined in such a way as it was initially merely thought (merely meant). ([32]: part 2, §8)[47]

This experience is what Husserl calls *fulfillment*. It is an act in its own right since it has a particular intentional correlate: the identity of the intentional object of the signitive act and the intentional object of the associated intuitive act (loc. cit.). According to this theory *cognition* is an act complex involving a higher-order act of identification which is *founded* on lower-order signitive and intuitive acts.[48]

For the purpose of illustration consider how the existence of the transitive closure of a set is recognized. For a given set a, we *define* its transitive closure as the smallest transitive set containing a as a subset.[49] The act A_1 corresponding to this definition is merely signitive since it fails to 'produce' an object *in propria persona*. Indeed, at least initially, it remains unclear whether there are any transitive sets containing a as a subset. In a separate act A_2 the intuitions underlying the axioms of union, infinity and replacement deliver an object, namely the set $a \cup \bigcup a \cup \bigcup \bigcup a \ldots$, which is transitive and contains a as a subset. Finally, an act of synthesis A_3 brings to consciousness the identity of the intentional object of A_1 with the intentional object of A_2.

Husserl's strategy for reconciling the subjectivity of the act of knowing with the objectivity of the content that is known exploits the idea of fulfillment in order to

[46] See §§1–29 and §§36–39 of the Sixth Logical Investigation in [32].

[47] Wir erleben es, wie in der Anschauung *dasselbe* Gegenständliche intuitive vergegenwärtigt ist, welches im symbolischen Akte "bloss gedacht" war, und dass es gerade als das so und so Bestimmte anschaulich wird, als was es zunächst bloss gedacht (bloss bedeutet) war.

[48] The concept of foundation with regard to acts and their components is crucial in Husserl's account of knowledge. Formally, "a content of the species α is founded in a content of the species β if an α by its essence ... cannot exist, unless also a β exists; where it is left open whether also the co-existence of certain γ, δ is required or not" ([32]: Investigation 3, §21). Higher-order acts are acts that are founded upon other acts in this sense.

[49] A set a is *transitive* if it is downward closed under the $\in$-relation, i.e., if every element of a is also a subset of a.

accommodate the traditional view that knowledge requires the *adaequatio rei ac intellectus*.

> [W]here a presentative intention has achieved its ultimate fulfillment in ideal and perfect perception, there the genuine *adaequatio rei et intellectus* has established itself: the objective is *really "present"* or *"given"* exactly as that which is intended; no longer is any partial intention implied that lacks its fulfillment. ([32]: part 2, §37)[50]

According to Husserl here the *intellectus* is the meaning intention of a signitive act, and the *adaequatio* is realized if an objectivity is explicitly given in intuition[51] exactly as it was meant in the signitive act. At the same time he is forced to concede that ultimate fulfillment by means of perception remains an ideal possibility.

> The object is not really given [in perception], for it is not fully and wholly given as what it itself is. It appears merely "from the front", only "perspectively shortened and foreshadowed" etc. ([32]: part 2, §14b)[52]

Consequently the purported *adaequatio* seems to be confined to the subjective sphere of the perceiver.[53] Husserl's answer to this dilemma appeals to the horizon structure of experience. Although an object of perception always appears under some aspect and is only one-sidedly present in consciousness, that presence includes adumbrations to other possible orientations in which it might be perceived. In this way misperceptions may be exposed eventually. The decisive point is that the horizon of potential future experiences is necessarily open-ended.

> No perception of [a] thing is final, there is always room for new perceptions that would determine indeterminations more accurately and fulfill what is unfulfilled. With every continuation the sum of the determinations of the thing-noema, which continuously belongs to the same thing X, enriches itself. It is a law of essence, that *every* perception and perception manifold is expandable, whence the process is unbounded; accordingly no intuitive grasping of the nature of a thing can be so complete that a further perception could not contribute to it something noematically new. ([33]: §149)[54]

---

[50] [W]o sich eine Vorstellungsintention durch diese ideal vollkommene Wahrnehmung letzte Erfüllung verschafft hat, da hat sich die echte *adaequatio rei et intellectus* hergestellt: das Gegenständliche ist genau als das, als welches es intendiert ist *wirklich "gegenwärtig"* oder *"gegeben"*; keine Partialintention ist mehr impliziert, die ihrer Erfüllung ermangele.

[51] Husserl uses the terms 'intuition' and 'perception' interchangeably in virtue of their identical roles in the fulfillment relation ([32]: part 2, §45).

[52] Der Gegenstand ist nicht wirklich gegeben, er ist nämlich nicht voll und ganz als derjenige gegeben, welcher er selbst ist. Er erscheint nur "von der Vorderseite", nur "perspektivisch verkürzt und abgeschattet" u. dgl.

[53] Optical illusions are a vivid illustration of this.

Interestingly, this yields a characterization of the *transcendence* of things, their being outside consciousness, in terms of the contents of consciousness. Objects, whether real or ideal,[55] have their determinate properties that go beyond what is experienced or anticipated, and these properties are there to be explored. In the course of this exploration we encounter objective constraints for example by way of physical boundary conditions or demands of logical consistency.[56] A perceptual noema is therefore always subject to more or less radical adjustments in the face of recalcitrant experiences or exposed misperceptions. This distinguishes perception from hallucination and other modes of consciousness with explicit awareness of an object. Husserl maintains that even though the resolution of ambiguous and indeterminate aspects of a given perception can proceed in infinitely many directions, it nevertheless proceeds systematically "in a strictly determinate fashion" ([33]: §44).[57] As a consequence the noema must furnish a rule for organizing the manifold of incoming perceptual data, recollections of past experiences, anticipations of and adumbrations to possible other experiences as well as the intrinsic relationships in which they stand to one another in such a way that they can be synthesized into the coherent experience of one and the same object. Thus the meaning of 'objective existence' consists in the progressive 'filling' of an open-ended system of 'unfilled' noematic components that correspond to possible perspectival variations of the object.

The reader will have noticed that this rough outline of Husserl's theory of perception and its epistemic role fails to address traditional metaphysical concerns. Clearly the purpose of phenomenology is not to deliver a philosophical 'proof' of the existence of the natural world or a mind-independent realm of mathematical objects. Rather, it aims to lay open the sense of those claims and the meaning of the evidence in their favor by consulting the intentional structures of consciousness. The reason why this is feasible in the first place is that all meaning, including what it means for an object to be outside of consciousness, is constituted in and dependent on consciousness ([33]: §44, §52, [34]: §94, 233).

10. If we agree that questions pertaining to the nature of axioms, their plausibility and the rationale behind their selection, are linked to the meanings of the terms they contain, and if we accept Husserl's thesis that meaning goes hand in hand with intentionality,

[54] Keine Wahrnehmung des Dinges ist letztabgeschlossen, immer bleibt Raum für neue Wahrnehmungen, die Unbestimmtheiten näher bestimmen, Unerfülltheiten erfüllen würden. Mit jedem Fortgange bereichert sich der Bestimmungsgehalt des Dingnoemas, das stetig zu demselben Dinge X gehört. Es ist eine Wesenseinsicht, dass jede Wahrnehmung und Wahrnehmungsmannigfaltigkeit erweiterungsfähig, der Prozess also ein endloser ist; demgemäss kann keine intuitive Erfassung des Dingwesens so vollständig sein, dass eine weitere Wahrnehmung ihr nicht noematisch Neues beifügen könnte.

[55] The first uncountable cardinal and its unresolved relation to the size of the continuum is a good example.

[56] The objection that the demand of logical consistency is *our* choice is quickly dismissed because if that were so, then we could also choose to bring about consistency *and* certain theorems ([21]: fn. 22, 314).

[57] According to Husserl this strictly determinate fashion is predelineated by the category to which the perceptual object belongs. In the case of spatial objects, e.g., one is bound by the laws of geometry. Cf. §142 and §§149ff of [33].

the next question is what kind of acts are involved in the constitution of the concept of set. The answer is straightforward and can be gleaned from Cantor's famous definition of 'set'.

> By a "set" we understand any collection M of definite well-distinguished objects m of our intuition or our thinking (which are called the "elements" of M) into a whole. ([8]: 282)[58]

Husserl, who was a colleague of Cantor's at the University of Halle from 1886 to 1900, had already analyzed the concept of set along these lines in his *Habilitationsschrift* of 1887 [31].[59] He actually distinguished two stages in the constitution of sets. In the first stage disjoint items are collected into a plurality that is a 'Many' not yet apprehended as a 'One'. At this point the set is preconstituted but it is not thematic as an object. In the second stage, by reflecting on the preceding act of collection, the unity of this plurality is brought to awareness, and we become conscious of the set as a new object.

> *A [set] arises when a uniform interest and simultaneously in it and with it a uniform awareness highlights distinct contents individually and contains them.* Thus the collective unity can only be comprehended by reflecting on the mental act giving rise to the set. ([31]: 74)[60]

Here and elsewhere[61] the reference to mental acts should not be misconstrued as saying that sets are somehow created in the mind of the mathematician. In other words, Husserl's analysis is not concerned with metaphysical questions about the conditions for the existence of sets. Rather he aims to show how our awareness of sets as objects is enabled. Thus there is no conflict with a platonistic interpretation of set theory; indeed, as noted earlier, mathematical objects are experienced as mind-independent.

The constitution of the concept of set requires a further act of abstraction in which the "general form" of all sets is singled out ([32]: part 2, §44). The adequacy of this analysis is apparent from the fact that it allows to 'deduce' the ZF axioms (i.e., ZFC without the axiom of choice) from the concept of set in its intended meaning. In those 'deductions' one is identifying the mental acts through which we become conscious

[58] Unter einer "Menge" verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die "Elemente" von M genannt werden) zu einem Ganzen.

[59] It is worth noting that Cantor's earlier definition of 'set' as a "Many which can be thought of as One" ([6]: 204), which predates Husserl's arrival in Halle by three years, does not explicitly refer to an act of collecting and has a more metaphysical flavor with its appeal to 'possibility'. Moreover, Cantor's 1895 definition appears almost *verbatim* in the opening paragraph of *Zur Lehre vom Inbegriff*, a manuscript prepared by Husserl in 1891 (cf. the Husserliana edition of [31]: 384).

[60] *Ein Inbegriff entsteht, indem ein einheitliches Interesse und in ihm und mit ihm zugleich ein einheitliches Bemerken verschiedene Inhalte für sich heraushebt und umfasst.* Es kann also die kollektive Verbindung auch nur erfasst werden durch Reflexion auf den psychischen Akt, durch welchen der Inbegriff zustande kommt.

[61] The constitutional analysis of sets in terms of a two-tiered process that involves an act of collecting followed by an act of reflection is fully upheld in §61 of the posthumously published [35].

of the sets in question. Husserl essentially carried this out for the pairing, union and separation axioms in the *Philosophie der Arithmetik*. To be sure, he did not yet have the idea of fulfillment then, and he is only talking about finite sets of sensory objects most of the time. But his arguments can be adapted to handle infinite sets by means of what he calls *figural moments*: directly perceivable *Gestalt-qualities* of symbolically presented multiplicities which are indicative of set character ([31]: ch. XI).[62] The details of this are worked out for the replacement axiom in [27]: Sec. 12.[63] It is worth emphasizing that the usual informal arguments for the ZF axioms found in textbooks of set theory *de facto* attempt to give a phenomenological justification for the axioms. Invariably these attempts, however, fall short of the latter in terms of transparency.

Similar remarks apply with respect to large cardinal axioms derived from reflection principles. The informal discussion of the iterative concept of set and the cumulative hierarchy in Section 2 can be recast as an analysis of the meaning of the expression V in terms of mental acts of iterated collections and their idealizations. These acts are linked into a complex network by the various kinds of dependence relations obtaining between them. One way in which acts can be related is via the filling of unfilled noematic components. For example, the intuitive act in which the progression of the ordinals is brought to consciousness fills a noematic component of the idealized notion '*arbitrary* iteration of the operation set of'.[64] In addition, the foundation relation between acts induces a hierarchical structure on the network.[65] In the first instance this concerns acts of abstraction such as the setting into relief of a dependent moment in a composite whole which is subsequently studied in its own right. Finally there are also *logical* relations between the noemata of various acts. For example, the noema of the act expressing the idea 'arbitrary iteration of the operation set of' is logically related to the noema of the act in which the absolute indescribability of V is grasped. Such relations introduce further dependences between acts of the sort 'a rational subject who performs act A_1 must also perform act A_2'.

By means of these kinds of dependence relations, i.e., in terms of the instrinsic structure of this network of acts, the meaning of the axiom claiming the existence of inaccessible cardinals in V can be explicated. Among other things, this makes use of the fact that the meaning intention expressed in the definition of an inaccessible cardinal is an instantiation of the type given by the meaning intention expressed in

[62] Ironically Husserl, worried about the logical impossibility of infinite sets ([31]: 218ff), did not use figural moments to that end.

[63] The main application of figural moments in that paper is to pin down the difference in evidential character between the axiom of choice and the logically equivalent wellorder principle. There is also a phenomenological argument why the immediate plausibility of the axiom of choice falls short of providing evidence of its truth, see [27]: Sec. 12. Incidentally, the ideas underlying that argument may be used to 'derive' the power set axiom phenomenologically. It is, however, a separate and difficult issue to determine the sense of the maximality principle which is implicit in the formulation of that axiom, cf. fn. 64. Some light can be shed on a particular aspect of this when the related notion 'arbitrary subset' is considered in the context of Husserl's remarks on *free variation* in [35]: §§87, 88.

[64] Another noematic component is filled by the intuition underlying the idea of the power set of a set M as the set of *all* subsets of M.

[65] For the concept of foundation see footnote 48.

Cantor's doctrine of the Absolute. Furthermore the fact that the acts in which the total indescribability of V is brought to awareness are appropriatly founded on the basic intuitions entering into Cantor's definition of 'set' allows to spell out the sense in which the axiom of inaccessible cardinals is a "natural continuation" ([20]: 182) of the standard axioms.

There is no analogous argument for measurable cardinals (cf. Section 3) when the meaning of V is construed along these lines. However, new components may be added to the noema of V, such as transcendence of V over Gödel's L. These components, although *prima facie* unrelated, are appropriately filled in the presence of a measurable cardinal. Given the meaning of 'objective existence' in terms of the progressive filling of new noematic components, and given our mathematical experience with measurable cardinals, we cannot help but postulating their existence. This is of course not a verification of the belief in their existence, but it enables us to understand that belief philosophically.

Projective determinacy is a principle whose formulation lacks an intuitive connection with the iterative concept of set. But the sense in which it provides a 'solution' to the problems of classical descriptive set theory is intimately linked to intuitive expectations contained in the meaning-horizon of '(projective) set'. From a geometric point of view projective sets as point sets in Euclidean space are conceived *as* 'simple' and therefore ought to have regularity properties such as Lebesgue measurability. All of these regularity properties provably obtain for the projective sets under PD. In virtue of this, PD is said to yield the 'correct' theory for the projective sets. This distinguishes descriptive set theory under PD from rival theories enjoying syntactical features such as uniformity in proofs and completeness.[66] But PD is itself a projective statement when formulated as a schema and thus belongs to this supposedly 'correct' theory. Therefore it was natural to look for a proof of PD from strong axioms of infinity. When that program was finally completed, the theory under PD turned out to be 'correct' in an even stronger sense (see the end of Section 5).

The point I wish to emphasize here is that in order fully to understand the sense in which the solution offered by PD can claim objectivity we must abandon the one-sided view that the objective is something entirely alien to the subjective and that it ought to be studied with complete disregard of the mental life of the mathematician. As illustrated by the previous section, already the intentionality inherent in the seemingly straightforward perception of physical objects seriously undermines a conception of objectivity derived from the Cartesian dualism of *res extensa* versus *res cogitans*. Likewise in the ascent to higher levels of objectivity an analysis aimed at uncovering the rationale behind the selection of first principles and the meanings of claims pertaining to their validity must incorporate the intentional structures of consciousness. To that end

[66] Strictly speaking completeness is not a syntactic notion because it refers to the decidability of 'natural' statements. The latter is obviously not a syntactic notion, cf. fn. 19 and 23.

> it is not enough that logic, in the manner characteristic of positive sciences, methodically fashion objective theories and reduce the forms of possible genuine theory to principles and norms. We must rise above the self-forgetfulness of the theorizer who, in his theoretical producing, devotes himself to the subject matter, the theories and the methods, and accordingly knows nothing of the inwardness of that producing—who lives in producing, but does not have this productive living itself as a theme within his field of vision. Only by virtue of a fundamental clarification, penetrating the depths of the inwardness that produces cognition and theory, the *transcendental* inwardness, does what is produced as genuine theory and genuine science become understandable. ([34]: 14, 15f in the translation by D. Cairns)[67]

The need for such a reorientation is particularly evident in the current situation with the continuum problem where there is some consensus among practitioners of set theory that any solution would have to be accompanied by an analysis of what it means to be a solution. Woodin's recent work, which has added an intriguing twist to this ongoing debate, serves as a case in point. In order to recover the sense in which his results provide evidence for the *falsity* of CH it is clearly not sufficient to study mathematical facts and their formal relation with various abstractions and idealizations. Rather one must examine how the mathematician is intentionally related to those facts because it is on these grounds that *he* accepts certain abstractions and idealizations as 'natural' or 'reasonable'.

As one among many examples consider the theorem about the definability of the set $O^{(\Omega)}$ in the structure $\langle H(c^+), \in\rangle$.[68] On the basis of the mathematical content, there is no question that $O^{(\Omega)}$ is a formal analog of O' and that the application of Tarski's theorem for obtaining non-CH resembles the formal structure of the heuristic argument for Gödel's first incompleteness theorem in [17]: 362f. Moreover, formal analogies of that sort suggest that Ω-logic is a non-arbitrary generalization of first order logic. However, this in itself does not tell us why the finite axiomatizability of the theory of $\langle H(\omega_2), \in\rangle$ should be regarded as a criterion for determining the *true* value of the continuum in the first place. Clearly Woodin defends the realistic attitude towards CH, in other words he aims at a resolution in V.[69] And it is equally clear that his approach involves a re-interpretation of the noema of V conceived no

[67] [kann es nicht damit sein Bewenden haben], dass die Logik in der Weise der positiven Wissenschaften objektive Theorien methodisch gestalte, und die Formen möglicher echter Theorie auf Prinzipien und Normen bringe. Wir müssen uns über die Selbstvergessenheit des Theoretikers erheben, der im theoretischen Leisten den Sachen, den Theorien und Methoden hingegeben, von der Innerlichkeit seines Leistens nichts weiss, der in ihnen lebt, aber dieses leistende Leben selbst nicht im thematischen Blick hat. Nur durch eine prinzipielle Klärung, die in die Tiefen der Erkenntnis und Theorie leistenden Innerlichkeit, der *transzendentalen* Innerlichkeit, hinabsteigt, wird, was als echte Theorie und echte Wissenschaft geleistet ist, verständlich.

[68] See [59]: 688. This uses the existence of a proper class of Woodin cardinals as a background assumption.

[69] This is also implicit in his more cautious claim that these results—no matter whether they are the solution—provide convincing evidence that there *is* a solution to the continuum problem ([59]: 690).

longer as encompassing all possible acts of collection, but rather as an epistemological attitude, the broadest possible point of view. In my opinion, the conceptual resources of phenomenology with its nuanced distinctions will be crucial in spelling out what exactly this means.

References

[1] Benacerraf, Paul and Hilary Putnam (eds.): 1983. *Philosophy of Mathematics: Selected Readings*, 2nd edition. Cambridge: Cambridge University Press.

[2] Bernays, Paul: 1961. Zur Frage der Unendlichkeitsschemata in der axiomatischen Mengenlehre. In: V. Bar-Hillel *et al*. (eds.), *Essays in the Foundations of Mathematics*, Jerusalem: Magnes Press, 3–49.

[3] Cantor, Georg: 1872. Über eine Ausdehnung des Satzes aus der Theorie der trigonometrischen Reihen. *Math. Annalen* 5: 123–132. Reprinted in [9]: 92–101.

[4] Cantor, Georg: 1874. Über eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen. *Crelles Journal f. Mathematik* 77: 258–262. Reprinted in [9]: 115–118.

[5] Cantor, Georg: 1878. Ein Beitrag zur Mannigfaltigkeitslehre. *Crelles Journal f. Mathematik* 84: 242–258. Reprinted in [9]: 119–133.

[6] Cantor, Georg: 1883. Grundlagen einer allgemeinen Mannigfaltigkeitslehre. *Math. Annalen* 21: 545–586. Reprinted in [9]: 165–208.

[7] Cantor, Georg: 1887. Mitteilungen zur Lehre vom Transfiniten. *Zeitschr. f. Philosophie und philos. Kritik* 91: 81–125. Reprinted in [9]: 378–419.

[8] Cantor, Georg: 1895, 1897. Beiträge zur Begründung der transfiniten Mengenlehre. *Math. Annalen* 46: 481–551 and 49: 207–246. Reprinted in [9]: 282–351.

[9] Cantor, Georg: 1932. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Edited by E. Zermelo. Berlin: Springer. Page numbers refer to the reprographic reproduction, Hildesheim: Georg Olms Verlagsbuchhandlung (1962).

[10] Cohen, Paul J.: 1963, 1964. *The Independence of the Continuum Hypothesis I & II*. Proc. Natl. Acad. Sci., USA 50 & 51: 1143–1148, 105–110.

[11] Dales, H. Garth and W. Hugh Woodin: 1987. *An Introduction to Independence for Analysts*. London Math. Soc. Lecture Notes Series 115. Cambridge: Cambridge University Press.

[12] Føllesdal, Dagfinn: 1969. Husserl's notion of the noema. *J. of Philosophy* 66, 680–687.

[13] Fraenkel Abraham A., Yehoshua Bar-Hillel, and Azriel Lévy: 1973. *Foundations of Set Theory*, 2nd edition. Amsterdam: North-Holland.

[14] Friedman, Harvey: 1986. Necessary uses of abstract theory in finite mathematics. *Advances in Math.* 60 (1): 92–122.

[15] Friedman, Harvey: 1998. Finite functions and the necessary use of large cardinals. *Annals of Math.* 148 (3): 803–893.

[16] Friedman, Harvey: 2000. Normal mathematics will need new axioms. *Bull. Symb. Logic* 6: 434–446.

[17] Gödel Kurt: 1934. On undecidable propositions of formal mathematical systems. In: M. Davis (ed.), *The Undecidable*, Hewlett, NY: Raven Press (1965). Reprinted in: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume I*, New York: Oxford University Press, 1986, 346–371.

[18] Gödel, Kurt: 1938. *The Consistency of the Axiom of Choice and the Generalized Continuum Hypothesis*. Proc. Natl. Acad. Sci., USA 24: 556–557. Reprinted in: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume II*, New York: Oxford University Press, 1990, 26–27.

[19] Gödel, Kurt: 1944. Russell's mathematical logic. In: P. Schilpp (ed.), *The Philosophy of Bertrand Russell*, Library of living philosophers, vol. 5, Evanston: Northwestern University, 125–153. Reprinted in: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume II*, New York: Oxford University Press, 1990, 119–141.

[20] Gödel, Kurt: 1947. What is Cantor's continuum problem? *American Mathematical Monthly* 54: 515–525. Reprinted in: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume II*, New York: Oxford University Press, 1990, 176–187.

[21] Gödel, Kurt: 1995. Some basic theorems on the foundations of mathematics and their implications. In: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume III*, New York: Oxford University Press, 1995, 304–323.

[22] Gödel, Kurt: 1961/?. The modern development of the foundations of mathematics in the light of philosophy. Transcription of a shorthand draft, with translation by E. Köhler and H. Wang, revised by J. Dawson, C. Parsons, and W. Craig. In: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume III*, New York: Oxford University Press, 1995, 374–387.

[23] Gödel, Kurt: 1964. What is Cantor's continuum problem? Revised and expanded version of [20], in [1]: 258–273. Reprinted in: S. Feferman *et. al.* (eds.), *Kurt Gödel Collected Works, Volume II*, New York: Oxford University Press, 1990, 254–270.

[24] Hauser, Kai: 1991. Indescribable cardinals and elementary embeddings. *J. Symb. Logic* 56: 439–457.

[25] Hauser, Kai: 2001. Objectivity over objects: A case study in theory formation. *Synthese* 128 (3): 245–285.

[26] Hauser, Kai: Gödel's program revisited. To appear in *Bull. Symb. Logic*.

[27] Hauser, Kai: Is choice self-evident? Submitted for publication.

[28] Hilbert, David: 1900. Mathematische Probleme. *Göttinger Nachrichten*: 253–257. English translation in *Proceedings of Symposia in Pure Mathematics* 28, Providence, RI: Amer. Math. Soc., 1976, 1–34.

[29] Hilbert, David: 1918. Axiomatisches Denken. *Math. Annalen* 78: 405–415.

[30] Hilbert, David: 1925. Über das Unendliche. *Math. Annalen* 95: 161–190.

[31] Husserl, Edmund: 1891. Philosophie der Arithmetik. Logische und psychologische Untersuchungen. Halle a. d. S.: C. E. M. Pfeffer (Robert Stricker). Reprinted in *Husserliana*

XII, The Hague: Martinus Nijhoff, 1970.

[32] Husserl, Edmund: 1913, 1968. *Logische Untersuchungen, Zweiter Band. I. Teil: Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*. Second Edition, Halle: Niemeyer. *II. Teil: Elemente einer phänomenologischen Aufklärung der Erkenntnis*. Fourth Edition, Tübingen: Niemeyer.

[33] Husserl, Edmund: 1928. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Erstes Buch*. Third Reprinting, Halle: Niemeyer.

[34] Husserl, Edmund: 1929. Formale und transzendentale Logik. Versuch einer Kritik der logischen Vernunft. *Jahrbuch für Philosophie und phänomenologische Forschung* 10: 1–298. English translation by D. Cairns, The Hague: Nijhoff, 1969.

[35] Husserl, Edmund: 1964. *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, Hamburg: Claassen, 1964.

[36] Husserl, Edmund: 1977. *Cartesianische Meditationen*, Hamburg: Felix Meiner.

[37] Jech, Thomas: 1978. *Set Theory*. New York: Academic Press.

[38] Kanamori, Akihiro: 1994. *The Higher Infinite*. Berlin: Springer.

[39] Kechris, Alexander S.: 2001. *Actions of Polish Groups and Classification Problems, Analysis and Logic*. London Math. Soc. Lecture Note Series. Cambridge: Cambridge University Press.

[40] Kunen, Kenneth: 1971. Elementary embeddings and infinitary combinatorics. *J. Symb. Logic* 36: 407–413.

[41] Kunen, Kenneth: 1980. *Set Theory*. Amsterdam: North-Holland.

[42] Levy, Azriel: 1960. Axiom schemata of strong infinity in axiomatic set theory. *Pacific Journal of Mathematics* 10: 223–238.

[43] Levy, Azriel and Robert M. Solvay: 1967. Measurable cardinals and the continuum hypothesis. *Israel J. Math*. 5: 234–248.

[44] Mahlo, Paul: 1911. Über lineare transfinite Mengen. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse* 63: 187–225.

[45] Mahlo, Paul.: 1912, 1913. Zur Theorie und Anwendung der ρ_0-Zahlen I, II. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse* 64: 108–122, 65: 268–282.

[46] Martin, Donald A. and John R. Steel: 1989. A proof of projective determinacy. *J. Am. Math. Soc*. 2: 71–125.

[47] Moore, Gregory H.: 1982. *Zermelo's Axiom of Choice*. New York: Springer.

[48] Reinhardt, William N.: 1970. Ackermann's set theory equals ZF. *Annals of Mathematical Logic* 2: 189–249.

[49] Reinhardt, William N.: 1974. Remarks on reflection principles, large cardinals, and elementary embeddings. In: T. Jech (ed.), *Axiomatic Set Theory*, Proc. of Symposia in Pure Mathematics 13, part 2, Providence: American Math. Soc., 189–205.

[50] Rota, Gian Carlo: 1997. The pernicious influence of mathematics on philosophy. In: *Indiscrete Thoughts*, Boston: Birkhäuser.

[51] Shelah, Saharon: 1974. Infinite abelian groups, Whitehead problem and some constructions. *Israel J. Math.* 18: 243–256.

[52] Shelah, Saharon: 1984. Can you take Solovay's inaccessible away? *Israel J. Math.* 48: 1–47.

[53] Solovay, Robert M.: 1970. A model of set theory in which every set of reals is Lebesgue measurable. *Ann. of Math.* (2) 92: 1–56.

[54] Steel, John R.: 2000. Mathematics needs new axioms. *Bull. Symb. Logic* 6: 422–433.

[55] Tieszen, Richard: 1995. Husserl and the philosophy of mathematics. In: B. Smith and D. Smith (eds.), *Cambridge Compagnion to Husserl*, Cambridge: Cambridge University Press, 438–462.

[56] Todorcevic, Stevo: 1989. Partition problems in topology. *Contemporary Mathematics* 84, Providence, RI: American Mathematical Society.

[57] Wang, Hao: 1996. *A Logical Journey. From Gödel to Philosophy*. Cambridge, MA: MIT Press.

[58] Woodin, W. Hugh: 1999. *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*. Berlin, New York: de Gruyter.

[59] Woodin, W. Hugh: 2001. The continuum hypothesis. *Notices of the Amer. Math. Soc.*: 567–576 and 681–690.

[60] Woodin, W. Hugh., A. R. D. Mathias, and Kai Hauser: *The Axiom of Determinacy*. To appear in de Gruyter Ser. Log. Appl.

Technische Universität Berlin
Fachbereich 3
Sekr. MA 8-1
10623 Berlin
Germany

E-mail: hauser@math.tu-berlin.de

Iterating Σ Operations in Admissible Set Theory without Foundation: A Further Aspect of Metapredicative Mahlo

Gerhard Jäger and Dieter Probst

Abstract. In this article we study the theory $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ which (i) describes a recursively inaccessible universe, (ii) permits the iteration of Σ operations along the ordinals, (iii) does not comprise $\in$ induction, and (iv) restricts complete induction on the natural numbers to sets. It is shown that the proof-theoretic ordinal of $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ is the metapredicative Mahlo ordinal $\varphi\omega 00$.

1. Introduction

The theory KPi^0, introduced in [3], is a natural set theory whose proof-theoretic strength is characterized by the well-known ordinal Γ_0. It describes a set-theoretic universe above the natural numbers as urelements which is an admissible limit of admissibles, i.e., recursively inaccessible. However, KPi^0 is very weak with respect to induction principles: $\in$ induction is not available at all, and complete induction on the natural numbers is restricted to sets.

In this article we study the effect of extending KPi^0 by exploring the possibility of iterating Σ operations (on the universe) along ordinals. It will be shown that the resulting system, we call it $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$, has the proof-theoretic ordinal $\varphi\omega 00$ and thus the same proof-theoretic strength as the theory KPm^0 of [6], which plays an important role in connection with the concept of metapredicative Mahlo; see [2] for a discussion of this topic in a wider context.

There exists a close relationship between our $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ and the system of second order arithmetic for Σ^1_1 transfinite dependent choice which Rüede describes in [8, 9]. Actually, we will adapt Rüede's well-ordering proof in [9] to the theory $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ in order to establish $\varphi\omega 00$ as its lower proof-theoretic bound. A straightforward argument yields that $\varphi\omega 00$ is also the upper proof-theoretic bound of $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$.

The plan of this paper is as follows: In the next section we introduce the basic definitions and describe the theories KPi^0, $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ and KPm^0. Section 3 is

dedicated to delimiting the system $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ from above by embedding it into KPm^0. Section 4 deals with some important properties of KPi^0 and $\mathsf{KPi}^0 + (\Sigma\text{-TR})$, whereas in Section 5 we carry through a well-ordering proof in $\mathsf{KPi}^0 + (\Sigma\text{-TR})$.

2. The Theories KPi^0, $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ and KPm^0

Let $\mathcal{L}_1$ be some of the standard languages of first order arithmetic with variables $a, b, c, d, e, f, g, h, u, v, w, x, y, z, \ldots$ (possibly with subscripts), a constant 0 as well as function and relation symbols for all primitive recursive functions and relations. The theory KPi^0 is now formulated in the extension $\mathcal{L}^\star = \mathcal{L}_1(\in, \mathsf{N}, \mathsf{S}, \mathsf{Ad})$ of $\mathcal{L}_1$ by the membership relation symbol $\in$, the set constant N for the set of natural numbers and the unary relation symbols S and Ad for sets and admissible sets, respectively.

The *number terms* of $\mathcal{L}^\star$ are inductively generated from the variables, the constant 0 and the symbols for the primitive recursive functions; the *terms* $(r, s, t, r_1, s_1, t_1, \ldots)$ of $\mathcal{L}^\star$ are the number terms of $\mathcal{L}_1$ plus the set constant N. The formulas $(A, B, C, A_1, B_1, C_1, \ldots)$ of $\mathcal{L}^\star$ as well as the Δ_0, Σ, Π, Σ_n and Π_n formulas of $\mathcal{L}^\star$ are defined as usual. Equality between objects is not represented by a primitive symbol, but defined by

$$(s = t) := \begin{cases} (s \in \mathsf{N} \wedge t \in \mathsf{N} \wedge (s =_{\mathsf{N}} t)) \vee \\ (\mathsf{S}(s) \wedge \mathsf{S}(t) \wedge (\forall x \in s)(x \in t) \wedge (\forall x \in t)(x \in s)) \end{cases}$$

where $=_{\mathsf{N}}$ is the symbol for the primitive recursive equality on the natural numbers. The formula A^s is the result of replacing each unrestricted quantifier $(\exists x)(\ldots)$ and $(\forall x)(\ldots)$ in A by $(\exists x \in s)(\ldots)$ and $(\forall x \in s)(\ldots)$, respectively. In addition, we freely make use of all standard set-theoretic notations and write, for example, $\mathsf{Tran}(s)$ for the Δ_0 formula saying that s is a transitive set.

Since the axioms of KPi^0 do not comprise $\in$ induction, i.e., the principle of foundation with respect to $\in$, we build it directly into the notion of *ordinal*,

$$\mathsf{Wf}(a, \in) := \forall x(x \subset a \wedge x \neq \emptyset \rightarrow (\exists y \in x)(\forall z \in y)(z \notin x)),$$

$$\mathsf{Ord}(a) := \mathsf{Tran}(a) \wedge (\forall x \in a)\mathsf{Tran}(x) \wedge \mathsf{Wf}(a, \in).$$

Thus $\mathsf{Ord}(a)$ is a Π formula of $\mathcal{L}^\star$; in the following lower case Greek letters range over ordinals.

The theory KPi^0 is formulated in the language $\mathcal{L}^\star$; its logical axioms comprise the usual axioms of classical first order logic with equality. The non-logical axioms of KPi^0 can be divided into the following five groups.

I. Ontological axioms. We have for all function symbols $\mathcal{H}$ and relation symbols $\mathcal{R}$ of $\mathcal{L}_1$ and all axioms $A(\vec{u})$ of group III whose free variables belong to the list $\vec{u}$:

$$a \in \mathsf{N} \leftrightarrow \neg \mathsf{S}(a), \tag{1}$$

$$\vec{a} \in \mathsf{N} \rightarrow \mathcal{H}(\vec{a}) \in \mathsf{N}, \tag{2}$$

$$\mathcal{R}(\vec{a}) \rightarrow \vec{a} \in \mathsf{N}, \tag{3}$$

$$a \in b \rightarrow \mathsf{S}(b), \tag{4}$$

$$\mathsf{Ad}(a) \rightarrow (\mathsf{N} \in a \wedge \mathsf{Tran}(a)), \tag{5}$$

$$\mathsf{Ad}(a) \rightarrow (\forall \vec{x} \in a) A^a(\vec{x}), \tag{6}$$

$$\mathsf{Ad}(a) \wedge \mathsf{Ad}(b) \rightarrow a \in b \vee a = b \vee b \in a. \tag{7}$$

II. Number-theoretic axioms. We have for all axioms $A(\vec{u})$ of Peano arithmetic PA which are not instances of the schema of complete induction and whose free variables belong to the list $\vec{u}$:

$$\vec{a} \in \mathsf{N} \rightarrow A^{\mathsf{N}}(\vec{a}). \qquad \text{(Number theory)}$$

III. Kripke–Platek axioms. We have for all Δ_0 formulas $A(u)$ and $B(u, v)$ of $\mathcal{L}^\star$:

$$\exists x(a \in x \wedge b \in x), \qquad \text{(Pair)}$$

$$\exists x(a \subset x \wedge \mathsf{Tran}(x)), \qquad \text{(Tran)}$$

$$\exists y(\mathsf{S}(y) \wedge y = \{x \in a : A(x)\}), \qquad (\Delta_0\text{-Sep})$$

$$(\forall x \in a)\exists y B(x, y) \rightarrow \exists z(\forall x \in a)(\exists y \in z) B(x, y). \qquad (\Delta_0\text{-Col})$$

IV. Limit axiom. It is used to formalize the assertion that each set is an element of an admissible set, hence we claim:

$$\exists x(a \in x \wedge \mathsf{Ad}(x)). \qquad \text{(Lim)}$$

V. Complete induction on N. The only induction principle included in the axioms of KPi^0 is the following axiom of complete induction on the natural numbers for sets:

$$0 \in a \wedge (\forall x \in \mathsf{N})(x \in a \rightarrow x + 1 \in a) \rightarrow \mathsf{N} \subset a. \qquad (\text{S-I}_{\mathsf{N}})$$

The monograph [1] provides an excellent introduction into general admissible set theory. Theories of admissible sets without foundation, on the other hand, have been studied, in particular, in [3, 4]. It is shown there, among other things, that the proof-theoretic ordinal of KPi^0 is Γ_0.

In Section 5 we will also mention an auxiliary *basic set theory* BS^0. It is obtained from KPi^0 by simply dropping the schema of Δ_0 collection. In this article, however, we are primarily interested in the axiom schema (Σ-TR) about the iteration of Σ operations, added to KPi^0. To formulate this principle, we introduce, for each Σ

formula $D(\vec{u}, x, y, z)$ of $\mathcal{L}^\star$ with at most the variables $\vec{u}, x, y, z$ free, the formula

$$\mathsf{Hier}_D(\vec{a}, b, f) := \begin{cases} \mathsf{Ord}(b) \wedge \mathsf{Fun}(f) \wedge \mathsf{Dom}(f) = b \wedge \\ (\forall \xi \in b) D(\vec{a}, \xi, f\restriction\xi, f(\xi)). \end{cases}$$

$\mathsf{KPi}^0 + (\Sigma\text{-TR})$ is now defined to be the theory obtained from KPi^0 by adding the axiom

$$\mathsf{Ord}(\alpha) \wedge (\forall \xi < \alpha) \forall x \exists! y D(\vec{a}, \xi, x, y) \rightarrow \exists f\, \mathsf{Hier}_D(\vec{a}, \alpha, f) \qquad (\Sigma\text{-TR})$$

for all Σ formulas $D(\vec{u}, x, y, z)$ of $\mathcal{L}^\star$. This axiom says that the Σ operation defined by D (depending on the parameters $\vec{a}$) can be iterated up to α.

$\mathsf{KPi}^0 + (\Sigma\text{-TR})$ is an interesting theory which reveals a further aspect of metapredicative Mahlo. In the sequel we will show that this system is proof-theoretically equivalent to the theory KPm^0 and that it has proof-theoretic ordinal $\varphi\omega 00$.

The theory KPm^0, introduced in [6], is also formulated in the language $\mathcal{L}^\star$ and extends KPi^0 by the schema of Π_2 reflection on the admissibles,

$$A(\vec{a}) \rightarrow \exists x (\vec{a} \in x \wedge \mathsf{Ad}(x) \wedge A^x(\vec{a})) \qquad (\Pi_2\text{-Ref}^{\mathsf{Ad}})$$

for all Π_2 formulas $A(\vec{u})$ of $\mathcal{L}^\star$ with at most the variables $\vec{u}$ free. In [6] it is also shown that KPm^0, i.e., $\mathsf{KPi}^0 + (\Pi_2\text{-Ref}^{\mathsf{Ad}})$, is of the same proof-theoretic strength as a natural system of explicit mathematics.

3. Embedding $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ into KPm^0

The only purpose of this very short section is to prove that $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ can be directly embedded into KPm^0. Given a Σ formula $D(\vec{u}, x, y, z)$ of $\mathcal{L}^\star$, parameters $\vec{a}$ for which D defines a Σ operation and an ordinal α, the principle of Π_2 reflection on admissibles tells us that there must exist an admissible set d, containing the parameters $\vec{a}$ and α, so that this operation maps d into d; working inside d allows us to iterate this operation up to α.

Lemma 1. *For all Σ formulas $D(\vec{u}, x, y, z)$ of $\mathcal{L}^\star$ with at most the variables $\vec{u}, x, y, z$ free, we have that*

$$\mathsf{KPm}^0 \vdash \mathsf{Ord}(\alpha) \wedge (\forall \xi < \alpha) \forall x \exists! y D(\vec{a}, \xi, x, y) \rightarrow \exists f\, \mathsf{Hier}_D(\vec{a}, \alpha, f).$$

Proof. We work informally in KPm^0 and choose an ordinal α and parameters $\vec{a}$ so that

$$(\forall \xi < \alpha) \forall x \exists! y D(\vec{a}, \xi, x, y). \qquad (1)$$

By Π_2 reflection on the admissibles applied to (1) it follows that there exists an admissible set d which contains α and $\vec{a}$ as elements and satisfies

$$(\forall \xi < \alpha)(\forall x \in d)(\exists y \in d) D^d(\vec{a}, \xi, x, y). \qquad (2)$$

Furthermore, because of Σ persistence and (1), we can now conclude from assertion (2) that

$$(\forall \xi < \alpha)(\forall x \in d)(\exists y \in d) D(\vec{a}, \xi, x, y), \tag{3}$$

$$(\forall \xi < \alpha)(\forall x, y \in d)(D(\vec{a}, \xi, x, y) \leftrightarrow D^d(\vec{a}, \xi, x, y)). \tag{4}$$

Working within the admissible set d, a straightforward adaptation of the usual proof of Σ recursion yields

$$(\exists f \in d)(\mathsf{Fun}(f) \wedge \mathsf{Dom}(f) = \beta \wedge (\forall \xi < \beta) D^d(\vec{a}, \xi, f \restriction \xi, f(\xi))) \tag{5}$$

by transfinite induction for all ordinals $\beta \leq \alpha$. Due to (4), the assertion of our lemma follows immediately from (5). □

This lemma states that (Σ-TR) is provable in KPm^0. Consequently, our system $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ is a subtheory of KPm^0.

Theorem 2 (Embedding). *For all $\mathcal{L}^\star$ formulas A we have that*

$$\mathsf{KPi}^0 + (\Sigma\text{-TR}) \vdash A \quad \Longrightarrow \quad \mathsf{KPm}^0 \vdash A.$$

In view of this theorem and the results of Jäger and Strahm [6], we know that the ordinal $\varphi\omega 00$ is an upper bound for the proof-theoretic strength of $\mathsf{KPi}^0 + (\Sigma\text{-TR})$.

4. Basic Properties of the Theories KPi^0 and $\mathsf{KPi}^0 + (\Sigma\text{-TR})$

The foundation axiom is not available in KPi^0, and thus it is not ruled out in general that there are sets which contain themselves. Nevertheless it can be shown that this is not possible for admissibles.

Lemma 3. *In KPi^0 it can be proved that*

$$\mathsf{Ad}(a) \rightarrow a \notin a.$$

Proof. Suppose, to the contrary, that a is an admissible set which contains itself as an element. By Δ_0 separation within a this a then also contains the "Russell" set r,

$$r := \{x \in a : x \notin x\}.$$

Hence we have that $r \in r$ if and only if $r \notin r$. This is a contradiction, and our lemma is proved. □

Now we turn to a further property of KPi^0 which will be crucial for the well-ordering proof in the next section: Every class $\mathcal{C}$ of admissibles that is definable by a Δ_0 formula of $\mathcal{L}^\star$ using parameters $\vec{a} \in \bigcap \mathcal{C}$, has a least element.

If $\in$ induction (foundation) were available we could simply carry over the standard recursion-theoretic proof. In the present context, however, a different strategy has to be chosen.

Lemma 4. *Let $D(\vec{u}, v)$ be a Δ_0 formula of $\mathcal{L}^\star$ with at most the variables $\vec{u}, v$ free. Then KPi^0 proves that*

$$\exists x(D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)) \rightarrow$$
$$\exists y(y = \bigcap\{x : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)\} \wedge D(\vec{a}, y) \wedge \vec{a} \in y \wedge \mathsf{Ad}(y)).$$

Proof. We work informally in the theory KPi^0 and fix some arbitrary parameters $\vec{a}$. The assertion is obvious by the linearity of the admissibles if the class $\{x : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)\}$ contains only finitely many elements. Hence we may assume that there exist elements b, c, d, e with the properties

$$b \in c \in d \in e, \tag{1}$$

$$b, c, d, e \in \{x : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)\}. \tag{2}$$

Now we use Δ_0 separation in order to define the set

$$s_0 := \bigcap\{x \in e : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)\}.$$

Because of the linearity of Ad we obtain the following further properties of this set s_0:

$$s_0 = \bigcap\{x : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)\}, \tag{3}$$

$$s_0 = \bigcap\{x \in c : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x)\}, \tag{4}$$

$$s_0 \in d. \tag{5}$$

Assertion (5) follows from (4) by Δ_0 separation in d. In a next step the further set

$$s_1 := \bigcap\{x \in e : D(\vec{a}, x) \wedge \vec{a} \in x \wedge \mathsf{Ad}(x) \wedge s_0 \in x\}$$

is introduced. Given these sets s_0 and s_1, we convince ourselves that they are really different,

$$s_0 \neq s_1. \tag{6}$$

Assume, to the contrary, that $s_0 = s_1$ and employ Δ_0 separation once more for defining the Russell set

$$r := \{x \in s_0 : x \notin x\}.$$

Then $r \in u$ for each set u satisfying $\mathsf{Ad}(u)$ and $s_0 \in u$. This implies $r \in s_1$ and therefore, because of our assumption, also $r \in s_0$. Therefore we have

$$r \in r \leftrightarrow r \in s_0 \wedge r \notin r \leftrightarrow r \notin r.$$

This is a contradiction, and so (6) is proved. This assertion (6), however, implies that there exists a set t with the property

$$t \in e \ \wedge \ D(\vec{a}, t) \ \wedge \ \vec{a} \in t \ \wedge \ \mathsf{Ad}(t) \ \wedge \ s_0 \notin t. \tag{7}$$

It remains to show that $s_0 = t$. The inclusion $s_0 \subset t$ is obvious. In order to prove $t \subset s_0$, we pick an arbitrary $u \in d$ with $D(\vec{a}, u) \ \wedge \ \vec{a} \in u \ \wedge \ \mathsf{Ad}(u)$ and establish $t \subset u$. The linearity of Ad gives us

$$t \in u \ \vee \ t = u \ \vee \ u \in t. \tag{8}$$

In the cases of $t \in u$ and $t = u$, the property $t \subset u$ is clear. On the other hand, $u \in t$ would imply that

$$s_0 \ = \ \bigcap \{x \in u \cup \{u\} : D(\vec{a}, x) \ \wedge \ \vec{a} \in x \ \wedge \ \mathsf{Ad}(x)\} \ \in \ t,$$

contradicting the choice of t. Therefore, putting everything together, we know that $s_0 = t$, hence $D(\vec{a}, s_0) \ \wedge \ \vec{a} \in s_0 \ \wedge \ \mathsf{Ad}(s_0)$. Remembering (3), this concludes the proof of our lemma. □

This lemma implies, for example, that in KPi^0 for any set a the intersection a^+ of all admissibles containing a is an admissible itself,

$$a^+ \ := \bigcap \{x : a \in x \ \wedge \ \mathsf{Ad}(x)\}.$$

Given a set a and a binary relation $b \subset a \times a$, we write $\mathsf{Lin}(a, b)$ if b is a strict linear ordering on a. A linear ordering is a well-ordering if any non-empty subset of its domain has a least element with respect to this ordering,

$$\mathsf{Wo}(a, b) := \mathsf{Lin}(a, b) \ \wedge \ \forall x (x \subset a \wedge x \neq \emptyset \ \rightarrow \ (\exists y \in x)(\forall z \in x)(\langle z, y\rangle \notin b)).$$

We first observe that in $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ all Σ operations can be iterated along arbitrary well-orderings, not only along ordinals as stated by $(\Sigma\text{-}\mathsf{TR})$. To do so, we write, for each Σ formula $D(\vec{u}, x, y, z)$ of $\mathcal{L}^\star$ with at most the variables $\vec{u}, x, y, z$ free, in analogy to Hier_D

$$\mathsf{Hier}_D^+(\vec{a}, b, c, f) := \begin{cases} \mathsf{Wo}(b, c) \ \wedge \ \mathsf{Fun}(f) \ \wedge \ \mathsf{Dom}(f) = b \ \wedge \\ (\forall x \in b) D(\vec{a}, x, \{\langle y, f(y)\rangle : \langle y, x\rangle \in c\}, f(x)). \end{cases}$$

For the proof of Theorem 6 below it is convenient to convince ourselves that for every well-ordering b on a set a there exists a function f—provable in KPi^0—so that the range $\mathsf{Rng}(f)$ of f is an ordinal and f is a 1-1 mapping from a to $\mathsf{Rng}(f)$ translating the order relation b on a into the $<$-relation on $\mathsf{Rng}(f)$. This function f is called the *collapse* of b onto a;

$$\mathsf{Clp}(a, b, f) := \begin{cases} b \subset a \times a \ \wedge \ \mathsf{Fun}(f) \ \wedge \ \mathsf{Dom}(f) = a \ \wedge \\ (\forall x \in a)(f(x) = \{f(y) : \langle y, x\rangle \in b\}). \end{cases}$$

The following lemma states that any well-ordering b on any set a has a collapse which is uniquely determined by a and b.

Lemma 5. *The following two assertions can be proved in* KPi^0*:*

1. $\mathsf{Wo}(a,b) \wedge \mathsf{Clp}(a,b,f) \wedge \mathsf{Clp}(a,b,g) \rightarrow f = g.$
2. $\mathsf{Wo}(a,b) \wedge a, b \in d \wedge \mathsf{Ad}(d) \rightarrow (\exists f \in d)\mathsf{Clp}(a,b,f).$

Proof. The uniqueness property is obtained by straightforward induction along the well-ordering b on a. For the proof of the existence of a collapse, we work informally in KPi^0 and choose any a, b, d so that

$$\mathsf{Wo}(a,b) \wedge a, b \in d \wedge \mathsf{Ad}(d).$$

Writing $a|u$ and $b||u$ for the restrictions of a and b to the predecessors of u,

$$a|u := \{x \in a : \langle x, u\rangle \in b\} \quad \text{and} \quad b||u := \{\langle x, y\rangle \in b : \langle y, u\rangle \in b\},$$

we can easily establish by induction along b on a that

$$(\exists! g \in d)\mathsf{Clp}(a|u, b||u, g)$$

for all elements u of a. With Δ_0 collection we can now immediately derive what we wish. □

Theorem 6. *For all* Σ *formulas* $D(\vec{u}, x, y, z)$ *of* $\mathcal{L}^\star$ *with at most the variables* $\vec{u}$*,* x*,* y*,* z *free, we have that* $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ *proves*

$$\mathsf{Wo}(b,c) \wedge (\forall x \in b)\forall y \exists! z D(\vec{a}, x, y, z) \rightarrow \exists f\, \mathsf{Hier}^+_D(\vec{a}, b, c, f)).$$

Proof. We work informally in $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ and choose arbitrary parameters $\vec{a}$ and sets b, c so that

$$\mathsf{Wo}(b,c) \wedge (\forall x \in b)\forall y \exists! z D(\vec{a}, x, y, z). \tag{1}$$

By the previous lemma there exists a collapse h of c onto b, i.e., we may assume that

$$\mathsf{Clp}(b,c,h) \tag{2}$$

for a suitable function h. Now set $\alpha := \mathsf{Rng}(h)$ and apply the principle $(\Sigma\text{-TR})$ to a properly tailored modification D' of the formula D. This gives us a function g which can then be modified to the desired witness for Hier_D. More precisely, let $B(u,v)$ be the Σ formula of $\mathcal{L}^\star$ which is the disjunction of the following three (mutually exclusive) $\mathcal{L}^\star$ formulas:

(i) $\mathsf{Ord}(u) \wedge v \in b \wedge h(v) = u,$

(ii) $u = \emptyset \wedge v = \emptyset,$

(iii) $\neg\mathsf{Ord}(u) \wedge (\forall x \in u)(\exists y \in b)\exists z(x = \langle h(y), z\rangle \wedge \langle y, z\rangle \in v) \wedge$
$(\forall x \in v)(\exists y \in b)\exists z(x = \langle y, z\rangle \wedge \langle h(y), z\rangle \in u).$

In addition we set

$$D'(\vec{a}, u, v, w) := \exists x \exists y (B(u, x) \wedge B(v, y) \wedge D(\vec{a}, x, y, w))$$

and observe that $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ proves

$$(\forall \xi < \alpha) \forall x \exists! y D'(\vec{a}, \xi, x, y). \tag{3}$$

In virtue of Σ reflection, the formula $D'(\vec{a}, u, v, w)$ is provably equivalent in $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ to a Σ formula. In view of (3) we can apply $(\Sigma\text{-}\mathsf{TR})$, and thus there exists a function g for which we have

$$\mathsf{Hier}_{D'}(\vec{a}, \alpha, g). \tag{4}$$

To finish our proof, let f be the function with domain b which is defined for all elements u of b by

$$f(u) := g(h(u)).$$

From (4), the definition of the formula $D'(\vec{a}, u, v, w)$ and this definition of f, we obtain

$$(\forall \xi < \alpha) \exists y \exists z (B(\xi, y) \wedge B(g \restriction \xi, z) \wedge D(\vec{a}, y, z, g(\xi))). \tag{5}$$

Because of (2) this immediately implies that

$$(\forall x \in b) \exists y \exists z (B(h(x), y) \wedge B(g \restriction h(x), z) \wedge D(\vec{a}, y, z, g(h(x)))). \tag{6}$$

In view of the definitions of the function f and the formula $B(u, v)$ we can transform this assertion into

$$(\forall x \in b) \exists z (B(\{\langle h(y), f(y)\rangle : \langle y, x\rangle \in c\}, z) \wedge D(\vec{a}, x, z, f(x))). \tag{7}$$

Looking at the definition of $B(u, v)$ once more, we see that (7) can be simplified to

$$(\forall x \in b) D(\vec{a}, x, \{\langle y, f(y)\rangle : \langle y, x\rangle \in c\}, f(x)). \tag{8}$$

This means, however, that we have $\mathsf{Hier}^+_D(\vec{a}, b, c, f)$, and therefore the proof of our lemma is completed. □

In his well-ordering proof for second order arithmetic with Σ^1_1 transfinite dependent choice, Rüede often makes use of Π^1_2 reflection on ω-models of ACA_0. In our present context, this part is taken over by Π_2 reflection on $\overline{\mathsf{Ad}}$. The "topological closure" $\overline{\mathsf{Ad}}$ of the predicate Ad is obtained by adding the limits of admissibles to the admissibles,

$$\overline{\mathsf{Ad}}(d) \;:=\; \mathsf{Ad}(d) \;\vee\; (d \neq \emptyset \wedge d = \bigcup \{x \in d : \mathsf{Ad}(x)\}).$$

Clearly each element of $\overline{\mathsf{Ad}}$ satisfies Δ_0 separation and models the theory BS^0. We prove a uniform version of Π_2 reflection on $\overline{\mathsf{Ad}}$ in a form tailored for our later purposes.

Lemma 7 (Π_2 reflection on $\overline{\mathsf{Ad}}$). *For any Σ formula $A(\vec{u}, v, w)$ of $\mathcal{L}^\star$ with at most the variables $\vec{u}, v, w$ free, there exists a Σ formula $A^\sharp(\vec{u}, v)$ of $\mathcal{L}^\star$ with at most the variables $\vec{u}, v$ free, so that the following two assertions can be proved in $\mathsf{KPi}^0+(\Sigma\text{-}\mathsf{TR})$:*

1. $\forall x \exists y A(\vec{a}, x, y) \rightarrow \exists! z A^\sharp(\vec{a}, z)$.
2. $\forall x \exists y A(\vec{a}, x, y) \rightarrow$
   $\forall z (A^\sharp(\vec{a}, z) \rightarrow \vec{a} \in z \wedge \overline{\mathsf{Ad}}(z) \wedge (\forall x \in z)(\exists y \in z) A^z(\vec{a}, x, y))$.

Proof. We work informally in $\mathsf{KPi}^0+(\Sigma\text{-}\mathsf{TR})$ and begin with introducing the following abbreviations:

$$B_A(\vec{u}, v, w) := \vec{u}, v \in w \wedge \mathsf{Ad}(w) \wedge (\forall x \in v)(\exists y \in w) A^w(\vec{u}, x, y),$$

$$C_A(\vec{u}, v, w) := B_A(\vec{u}, v, w) \wedge (\forall z \in w) \neg B_A(\vec{u}, v, z).$$

Obviously, $B_A(\vec{u}, v, w)$ and $C_A(\vec{u}, v, w)$ are Δ_0 formulas, and with Σ reflection and the limit axiom (Lim) we obtain for arbitrary parameters $\vec{a}$ that

$$\forall x \exists y A(\vec{a}, x, y) \rightarrow \forall v \exists w B_A(\vec{a}, v, w). \tag{1}$$

Hence Lemma 3, i.e., the fact that an admissible cannot contain itself, and Lemma 4 imply in view of (1) that

$$\forall x \exists y A(\vec{a}, x, y) \rightarrow \forall v \exists! w C_A(\vec{a}, v, w). \tag{2}$$

The operation described by this fact will now be iterated along the standard less relation $<_{\mathsf{N}}$ on the natural numbers N. To adjust everything to the formulation of Theorem 6 we define

$$D_A(\vec{u}, x, y, z) := C_A(\vec{u}, \bigcup\{v : (\exists w \in \mathsf{N})(w <_{\mathsf{N}} x \wedge \langle w, v\rangle \in y)\}, z).$$

Trivially, we have $\mathsf{Wo}(\mathsf{N}, <_{\mathsf{N}})$. Furthermore, (2) implies that

$$\forall x \exists y A(\vec{a}, x, y) \rightarrow (\forall x \in \mathsf{N}) \forall y \exists! z D_A(\vec{a}, x, y, z). \tag{3}$$

In virtue of Theorem 6 we consequently know that there exists a function f whose domain is the set N and which satisfies $\mathsf{Hier}^+_{D_A}(\vec{a}, \mathsf{N}, <_{\mathsf{N}}, f)$, i.e.,

$$\forall x \exists y A(\vec{a}, x, y) \rightarrow \exists f\, \mathsf{Hier}^+_{D_A}(\vec{a}, \mathsf{N}, <_{\mathsf{N}}, f). \tag{4}$$

The function f which is claimed to exist in (4) has to be unique. Thus also the set d,

$$d := \bigcup\{f(x) : x \in \mathsf{N}\},$$

is uniquely determined. Under the assumption $\forall x \exists y A(\vec{a}, x, y)$ it is now easily verified that we have for this set d the desired properties $\vec{a} \in d$, $\overline{\mathsf{Ad}}(d)$ and $(\forall x \in d)(\exists y \in d) A^d(\vec{a}, x, y))$.

To finish the proof of our lemma, we set

$$A^\sharp(\vec{u}, v) := \exists f (\mathsf{Hier}^+_{D_A}(\vec{u}, \mathsf{N}, <_{\mathsf{N}}, f) \wedge v = \bigcup\{f(x) : x \in \mathsf{N}\}).$$

Our previous considerations make it clear that for this Σ formula $A^\sharp(\vec{u}, v)$ both assertions of our lemma are satisfied. □

5. The Well-Ordering Proof in $\mathsf{KPi}^0 + (\Sigma\text{-TR})$

Now the stage is set for extending the well-ordering proof in [5] to our set theory $\mathsf{KPi}^0 + (\Sigma\text{-TR})$, similar to how Rüede [8] adopts it for the treatment of Σ^1_1 transfinite dependent choice. We will show that all ordinals less than $\varphi\omega 00$ are provable in $\mathsf{KPi}^0 + (\Sigma\text{-TR})$.

As in, for example, [6] we work with the *ternary Veblen functions* for coping with a sufficiently long initial segment of the ordinals. The usual Veblen hierarchy is generated by the binary function φ, starting off with $\varphi 0\beta = \omega^\beta$, and often discussed in the literature, cf. e.g., [7] or [10]. The ternary Veblen function φ is easily obtained from the binary φ as follows:

1. $\varphi 0\beta\gamma$ is $\varphi\beta\gamma$.

2. If $\alpha > 0$, then $\varphi\alpha 0\gamma$ denotes the γth ordinal which is strongly critical with respect to all functions $\lambda\xi.\lambda\eta.\varphi\delta\xi\eta$ for $\delta < \alpha$.

3. If $\alpha > 0$ and $\beta > 0$, then $\varphi\alpha\beta\gamma$ denotes the γth common fixed point of the functions $\lambda\eta.\varphi\alpha\delta\eta$ for $\delta < \beta$.

Let Ξ_0 be the least ordinal greater than 0 which is closed under addition and the ternary φ. In the following we will work with a standard primitive recursive notation system $(\mathsf{OT}, \prec)$ for all ordinals less than Ξ_0. All required definitions are straightforward generalizations of those used for building a notation system for Γ_0 (cf. [7, 10]) and can be omitted.

In this section we let $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \ldots$ (possibly with subscripts) range over the set OT; in addition, ℓ is used for codes of limit ordinals; the terms $\hat{0}, \hat{1}, \hat{2}, \ldots$ act as codes for the finite ordinals. To simplify the notation we often write the ordinal constants and ordinal functions such as,

$$0, \quad 1, \quad \omega, \quad \lambda\xi.\lambda\eta.(\xi + \eta), \quad \lambda\xi.\omega^\xi, \quad \lambda\zeta.\lambda\xi.\lambda\eta.\varphi\zeta\xi\eta$$

instead of the corresponding codes and primitive recursive functions. Another useful binary operation on ordinal notations, introduced in [5], is given by

$$\mathfrak{a} \uparrow \mathfrak{b} := \exists\mathfrak{c}\exists\ell(\mathfrak{b} = \mathfrak{c} + \mathfrak{a} \cdot \ell).$$

For completeness we also recall how it is expressed that our specific primitive recursive relation $\prec$ is a *well-ordering*, that a formula is *progressive* with respect to $\prec$ and how

transfinite induction along $\prec$ is defined for arbitrary formulas:

$$\mathsf{Wo}(\mathfrak{a}) := \mathsf{Wo}(\{\mathfrak{b} : \mathfrak{b} \prec \mathfrak{a}\}, \{\langle \mathfrak{c}, \mathfrak{b}\rangle : \mathfrak{c} \prec \mathfrak{b} \prec \mathfrak{a}\}),$$

$$\mathsf{Prog}(A) := \forall \mathfrak{a}((\forall \mathfrak{b} \prec \mathfrak{a}) A(\mathfrak{b}) \to A(\mathfrak{a})),$$

$$\mathsf{TI}(A, \mathfrak{a}) := \mathsf{Prog}(A) \to (\forall \mathfrak{b} \prec \mathfrak{a}) A(\mathfrak{b}).$$

Maybe apart from $\mathfrak{a} \uparrow \mathfrak{b}$, all these notions are standard in the context of well-ordering proofs. For dealing with $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ we need further predicates $\mathcal{K}_n(u)$ and $\mathcal{H}_n(\mathfrak{a}, u, f)$ which are defined simultaneously by induction on the natural number n as well as the predicates $\mathcal{I}(\mathfrak{b}, f, \mathfrak{a})$ and $\mathcal{M}_n(\mathfrak{b}, f, \mathfrak{a})$:

$$\mathcal{T}(f) := \mathsf{Fun}(f) \wedge \mathsf{Dom}(f) = \mathsf{OT},$$

$$\mathcal{K}_1(a) := \mathsf{Ad}(a),$$

$$\mathcal{K}_{n+1}(a) := \overline{\mathsf{Ad}}(a) \wedge [\forall x \exists f(\mathcal{T}(f) \wedge \forall \mathfrak{a}(\mathsf{Wo}(\mathfrak{a}) \to \mathcal{H}_n(\mathfrak{a}, x, f)))]^a,$$

$$\mathcal{H}_n(\mathfrak{a}, u, f) := \mathcal{T}(f) \wedge (\forall \mathfrak{b} \prec \mathfrak{a})(f\restriction \mathfrak{b} \in f(\mathfrak{b}) \wedge u \in f(\mathfrak{b}) \wedge \mathcal{K}_n(f(\mathfrak{b}))),$$

$$\mathcal{I}(\mathfrak{b}, f, \mathfrak{a}) := (\forall \mathfrak{c} \prec \mathfrak{b})(\forall x \in f(\mathfrak{c}))\mathsf{TI}(x, \mathfrak{a}),$$

$$\mathcal{M}_n(\mathfrak{b}, f, \mathfrak{a}) := \forall \mathfrak{c}(\forall \mathfrak{d} \preceq \mathfrak{b})(\omega^{1+\mathfrak{a}} \uparrow \mathfrak{d} \wedge \mathcal{I}(\mathfrak{d}, f, \mathfrak{c}) \to \mathcal{I}(\mathfrak{d}, f, \varphi \hat{n} \mathfrak{a} \mathfrak{c})).$$

The first lemma concerning this machinery states that each set a is provably an element of a set b which satisfies the property $\mathcal{K}_n$. It plays a key role in our well-ordering proof.

Lemma 8 (Main lemma). *For any natural number $n > 0$, there exists a Σ formula $F_n(u, v)$ of $\mathcal{L}^\star$ so that $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ proves*

$$\forall x \exists! y F_n(x, y) \wedge \forall x \forall y (F_n(x, y) \to x \in y \wedge \mathcal{K}_n(y)).$$

Proof. We prove this assertions by complete induction on n. For $n = 1$ we simply have to set

$$F_1(u, v) := (v = u^+).$$

Due to Lemma 4 and the discussion following this lemma we know that this formula $F_1(u, v)$ satisfies our requirements. Now we assume $n > 1$ and apply the induction hypothesis to provide a Σ formula $F_{n-1}(u, v)$ so that $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ proves

$$\forall x \exists! y F_{n-1}(x, y) \wedge \forall x \forall y (F_{n-1}(x, y) \to x \in y \wedge \mathcal{K}_n(y)). \tag{1}$$

Based on this Σ formula $F_{n-1}(u, v)$ we introduce the auxiliary Σ formula $B_n(u, v, w)$,

$$B_n(u, v, w) := F_{n-1}(\{u, v\}, w).$$

From (1) we immediately conclude that $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ proves

$$\forall x \forall y \exists! z B_n(x, y, z). \tag{2}$$

Hence $B_n(u, v, w)$ defines a Σ operation to which we want to apply Theorem 6 in a next step. This theorem implies for any parameter u and any element $\mathfrak{c}$ of OT that, provably in $\mathsf{KPi}^0 + (\Sigma\text{-TR})$,

$$\mathsf{Wo}(\mathfrak{c}) \;\rightarrow\; \begin{cases} \exists g[\mathsf{Fun}(g) \wedge \mathsf{Dom}(g) = \{\mathfrak{d} : \mathfrak{d} \prec \mathfrak{c}\} \wedge \\ \quad (\forall \mathfrak{d} \prec \mathfrak{c}) B_n(u, g\restriction \mathfrak{d}, g(\mathfrak{d}))]. \end{cases} \tag{3}$$

Now define $E_n(u, \mathfrak{c}, g)$ to be the Σ formula

$$\mathsf{Fun}(g) \wedge \mathsf{Dom}(g) = \{\mathfrak{d} : \mathfrak{d} \prec \mathfrak{c}\} \wedge (\forall \mathfrak{d} \prec \mathfrak{c}) B_n(u, g\restriction \mathfrak{d}, g(\mathfrak{d})).$$

Until the end of this section we work informally in $\mathsf{KPi}^0 + (\Sigma\text{-TR})$ and rewrite statement (3) as

$$\forall \mathfrak{c} \exists g(\mathsf{Wo}(\mathfrak{c}) \rightarrow E_n(u, \mathfrak{c}, g)). \tag{4}$$

Σ reflection, the limit axiom (Lim) and Σ persistence in connection with (4) guarantee the existence of an admissible set d such that

$$\forall \mathfrak{c}(\exists g \in d)(\mathsf{Wo}^d(\mathfrak{c}) \rightarrow E_n^d(u, \mathfrak{c}, g)). \tag{5}$$

We further claim that for all elements g and g' of d

$$\mathsf{Wo}^d(\mathfrak{c}) \wedge \mathfrak{a} \prec \mathfrak{b} \prec \mathfrak{c} \wedge E_n^d(u, \mathfrak{b}, g) \wedge E_n^d(u, \mathfrak{c}, g') \;\rightarrow\; g(\mathfrak{a}) = g'(\mathfrak{a}), \tag{6}$$

a fact that can be easily checked by inspecting the definitions of $E_n(u, \mathfrak{b}, g)$ and $E_n(u, \mathfrak{c}, g')$. A next step in the proof of our lemma is to set

$$f := \bigcup \{g \in d : \exists \mathfrak{c}(\mathsf{Wo}^d(\mathfrak{c}) \wedge E_n^d(u, \mathfrak{c}, g)\} \cup \{\langle \mathfrak{c}, \emptyset\rangle : \neg \mathsf{Wo}^d(\mathfrak{c})\}.$$

f is an element of d^+; by (6) we know that it is a function, and $\mathsf{Dom}(f) = \mathsf{OT}$ is immediate from its definition. Assertion (5) and the definition of f yield, in addition, that

$$\mathsf{Wo}^d(\mathfrak{c}) \;\rightarrow\; (\forall \mathfrak{d} \prec \mathfrak{c})(u \in f(\mathfrak{d}) \wedge f\restriction \mathfrak{d} \in f(\mathfrak{d}) \wedge \mathcal{K}_{n-1}(f(\mathfrak{d}))) \tag{7}$$

and this, in turn, implies because of Σ persistence and our previous remarks about f that

$$\forall x \exists f(\mathcal{T}(f) \wedge \forall \mathfrak{c}(\mathsf{Wo}(\mathfrak{c}) \rightarrow \mathcal{H}_{n-1}(\mathfrak{c}, x, f))). \tag{8}$$

The last step of our proof consists in applying Π_2 reflection on $\overline{\mathsf{Ad}}$, to the Σ formula $A_n(u, v, w)$,

$$A_n(u, v, w) := (u = u) \wedge \mathcal{T}(w) \wedge \forall \mathfrak{c}(\mathsf{Wo}(\mathfrak{c}) \rightarrow \mathcal{H}_{n-1}(\mathfrak{c}, v, w)).$$

Lemma 7 implies the existence of a Σ formula $A^{\sharp}_n(u, v)$ so that from (8) we may deduce for any parameter a that

$$\exists! y A^{\sharp}_n(a, y), \tag{9}$$

$$\forall y(A^{\sharp}_n(a, y) \ \rightarrow \ a \in y \ \wedge \ \overline{\mathsf{Ad}}(y) \ \wedge \ (\forall x \in y)(\exists f \in y)A^{y}_n(a, x, f)). \tag{10}$$

By choosing $F_n(u, v)$ to be the formula $A^{\sharp}_n(u, v)$, the assertions (9) and (10) together with the definition of the formula $\mathcal{K}_n(v)$ immediately imply

$$\forall x \exists! y F_n(x, y) \ \wedge \ \forall x \forall y(F_n(x, y) \ \rightarrow \ x \in y \ \wedge \ \mathcal{K}_n(y)).$$

This finishes the induction step, and therefore also the proof of our lemma is completed. □

The proofs of the following three lemmas can be easily recaptured by (more or less notational) adaptations of the corresponding proofs in [5] and [8]. Therefore we omit all details and confine ourselves to providing exact references.

Lemma 9. *The following three assertions can be proved in* BS^0*:*

1. $\mathcal{H}_1(\ell, u, f) \wedge \mathcal{I}(\ell, f, \mathfrak{a}) \ \rightarrow \ \mathcal{I}(\ell, f, \varphi\mathfrak{a}0)$.
2. $\mathcal{H}_1(\ell, u, f) \ \rightarrow \ \mathsf{Prog}(\{\mathfrak{a} : \mathcal{I}(\ell, f, \varphi 10\mathfrak{a})\})$.
3. $\mathcal{H}_1(\mathfrak{b}, u, f) \ \rightarrow \ \mathsf{Prog}(\{\mathfrak{a} : \mathcal{M}_1(\mathfrak{b}, f, \mathfrak{a})\}$.

Proof. For all details concerning the proof of these three assertions, see Lemma 5, Lemma 6 and Lemma 7 of [5]. □

Lemma 10. *For any natural number* $n > 0$*, the following three assertions can be proved in* BS^0*:*

1. $\mathcal{K}_{n+1}(a) \ \wedge \ [\forall x \forall f \forall \mathfrak{b}(\mathcal{H}_n(\mathfrak{b}, x, f) \rightarrow \mathsf{Prog}(\{\mathfrak{c} : \mathcal{M}_n(\mathfrak{b}, f, \mathfrak{c})\}))]^a$
   $\rightarrow \ \forall \mathfrak{c}[(\forall x \in a)\mathsf{TI}(x, \mathfrak{c}) \rightarrow (\forall x \in a)\mathsf{TI}(x, \varphi\hat{n}\mathfrak{c}0)]$.
2. $\mathcal{K}_{n+1}(a) \ \wedge \ \forall \mathfrak{c}[(\forall x \in a)\mathsf{TI}(x, \mathfrak{c}) \rightarrow (\forall x \in a)\mathsf{TI}(x, \varphi\hat{n}\mathfrak{c}0)]$
   $\rightarrow \ \mathsf{Prog}(\{\mathfrak{c} : (\forall x \in a)\mathsf{TI}(x, \varphi(\hat{n}{+}1)0\mathfrak{c})\})$.
3. $\mathcal{H}_n(\mathfrak{b}, u, f) \ \wedge \ \forall a[\mathcal{K}_n(a) \rightarrow \mathsf{Prog}(\{\mathfrak{c} : (\forall x \in a)\mathsf{TI}(x, \varphi\hat{n}0\mathfrak{c})\})]$
   $\rightarrow \ \mathsf{Prog}(\{\mathfrak{c} : \mathcal{M}_n(\mathfrak{b}, f, \mathfrak{c}\})$.

Proof. For all details concerning the proof of these three assertions see Lemma 4, Lemma 5 and Lemma 6 of [8]. □

Lemma 11. *For any natural number* $n > 0$*, the following three assertions can be proved in* BS^0*:*

1. $\overline{\mathsf{Ad}}(a) \rightarrow [\forall x \forall f \forall \mathfrak{b} \mathcal{H}_n(\mathfrak{b}, x, f) \rightarrow \mathsf{Prog}(\{\mathfrak{c} : \mathcal{M}_n(\mathfrak{b}, f, \mathfrak{c})\})]^a$.
2. $\mathcal{K}_{n+1}(a) \rightarrow \forall \mathfrak{c}[(\forall x \in a)\mathsf{TI}(x, \mathfrak{c}) \rightarrow (\forall x \in a)\mathsf{TI}(x, \varphi \hat{n} \mathfrak{c} 0)]$.
3. $\mathcal{K}_{n+1}(a) \rightarrow \mathsf{Prog}(\{\mathfrak{c} : (\forall x \in a)\mathsf{TI}(x, \varphi(\hat{n}{+}1)0\mathfrak{c})\})$.

Proof. Start showing that the first assertion implies the second and the second the third. Then prove the first assertion by induction on n. For details see Theorem 6 of [8]. □

Theorem 12 (Lower bound). *For any natural number n, we have that*

$$\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR}) \vdash \forall x \mathsf{TI}(x, \varphi \hat{n} 00).$$

Proof. Fix any natural number $n > 0$. Arguing informally in $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$, let a be an arbitrary set. In view of Lemma 8 we know that there exist a set b satisfying

$$a \in b \wedge \mathcal{K}_{n+1}(b). \tag{1}$$

The second part of Lemma 11 yields, in addition, that

$$\mathcal{K}_{n+1}(b) \rightarrow \forall \mathfrak{c}[(\forall x \in b)\mathsf{TI}(x, \mathfrak{c}) \rightarrow (\forall x \in b)\mathsf{TI}(x, \varphi \hat{n} \mathfrak{c} 0)], \tag{2}$$

and, consequently, if we set $\mathfrak{c} = 0$,

$$\mathcal{K}_{n+1}(b) \rightarrow (\forall x \in b)\mathsf{TI}(x, \varphi \hat{n} 00)]. \tag{3}$$

From (1) and (3) we conclude $\mathsf{TI}(a, \varphi \hat{n} 00)$, which is exactly what we had to show. □

An ordinal α is called provable in the theory T—formulated in $\mathcal{L}^\star$ or a similar language—if there exists a primitive recursive well-ordering $\lhd$ on the natural numbers of order-type α so that

$$T \vdash \mathsf{Wo}(\mathsf{N}, \lhd).$$

The least ordinal which is not provable in T is called the *proof-theoretic ordinal of* T and denoted by $|\mathsf{T}|$.

From Theorem 2, Theorem 12 above and the results of Jäger and Strahm [6], which tell us that $|\mathsf{KPm}^0| \leq \varphi\omega 00$, we derive the following characterization of the theory $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ in terms of their proof-theoretic ordinal.

Corollary 13. *The set theories* $\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})$ *and* KPm^0 *have the same proof-theoretic strength, namely*

$$|\mathsf{KPi}^0 + (\Sigma\text{-}\mathsf{TR})| = |\mathsf{KPm}^0| = \varphi\omega 00.$$

Of course, Theorem 2 and Theorem 12 also show that any ordinal less than $\varphi\omega 00$ is provable in KPm^0. A direct proof of this result can be found in Strahm [11].

Acknowledgement. Research partly supported by the Swiss National Science Foundation.

References

[1] Barwise, Jon: 1975. *Admissible Sets and Structures*. Berlin: Springer.

[2] Jäger, Gerhard: To appear. *Metapredicative and Explicit Mahlo: A Proof-Theoretic Perspective*. Logic Colloquium 2000.

[3] Jäger, Gerhard: 1984. The strength of admissibility without foundation. *The Journal of Symbolic Logic* 49 no. 3: 867–879.

[4] Jäger, Gerhard: 1986. *Theories for Admissible Sets: A Unifying Approach to Proof Theory*. Napoli: Bibliopolis.

[5] Jäger, Gerhard, Reinhard Kahle, Anton Setzer and Thomas Strahm: 1999. The proof-theoretic analysis of transfinitely iterated fixed point theories. *The Journal of Symbolic Logic* 64 no. 1: 53–67.

[6] Jäger, Gerhard and Thomas Strahm: 2001. Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory. *The Journal of Symbolic Logic* 66 no. 2: 935–958.

[7] Pohlers, Wolfram: 1989. *Proof Theory: An Introduction*. Lecture Notes in Mathematics, vol. 1407. Berlin: Springer.

[8] Rüede, Christian: 2000. *Metapredicative Subsystems of Analysis*. Ph.D. thesis. Institut für Informatik und angewandte Mathematik, Universität Bern.

[9] Rüede, Christian: 2003. The proof-theoretic analysis of Σ^1_1 transfinite dependent choice. *Annals of Pure and Applied Logic* 122 no. 1: 195–234.

[10] Schütte, Kurt: 1977. *Proof Theory*. Berlin: Springer.

[11] Strahm, Thomas: 2002. Wellordering proofs for metapredicative Mahlo. *The Journal of Symbolic Logic* 67 no. 1: 260–278.

Institut für Informatik und angewandte Mathematik
Universität Bern
Neubrückstrasse 10
3012 Bern
Switzerland

E-mail: jaeger@iam.unibe.ch
probst@iam.unibe.ch

Typical Ambiguity: Trying to Have Your Cake and Eat It Too

Solomon Feferman

Would ye both eat your cake and have your cake?

John Heywood, *Proverbs*[1]

Abstract. Ambiguity is a property of syntactic expressions which is ubiquitous in all informal languages—natural, scientific and mathematical; the efficient use of language depends to an exceptional extent on this feature. Disambiguation is the process of separating out the possible meanings of ambiguous expressions. Ambiguity is typical if the process of disambiguation can be carried out in some systematic way. Russell made use of typical ambiguity in the theory of types in order to combine the assurance of its (apparent) consistency ("having the cake") with the freedom of the informal untyped theory of classes and relations ("eating it too"). The paper begins with a brief tour of Russell's uses of typical ambiguity, including his treatment of the statement $Cls \in Cls$. This is generalized to a treatment in simple type theory of statements of the form $A \in B$ where A and B are class expressions for which A is prima facie of the same or higher type than B. In order to treat mathematically more interesting statements of self membership we then formulate a version of typical ambiguity for such statements in an extension of Zermelo–Fraenkel set theory. Specific attention is given to how the "naive" theory of categories can thereby be accounted for.

1. Ambiguity, Disambiguation and Typical Ambiguity

Ambiguity is a property of syntactic expressions which is ubiquitous in all informal (if not formal) languages—natural, scientific and mathematical; the efficient use of language depends to an exceptional extent on this feature. *Disambiguation* is the process of separating out the possible meanings of ambiguous expressions. Ambiguity is *typical*, or systematic, if the process of disambiguation can be carried out in some uniform way. In natural languages this is determined by specification of context. For example, the words *'I', 'you', 'here', 'now', 'this', 'my'*, and *'your'* used in daily

[1] Ascribed to John Heywood (c. 1497–1580) in [3] at Heywood 13.

language have meanings which vary with the context of utterance in a uniform way. Other words with more than one meaning such as 'bank', 'duck', and 'visiting', such as in the utterances 'He sat by the bank,' 'I saw her duck,' and 'Visiting relatives can be boring', can also be disambiguated from the context but not in any uniform way.[2]

Ambiguity in mathematical language involving such words as 'member,' 'union,' and 'complement,' is typical in that they are disambiguated in a uniform way relative to a context given by an underlying universe; similarly for the 'identity map.' In order to disambiguate arithmetic and algebraic symbols such as '0,' '1,' '$+$,' '$\times$,' and '$\leq$,' we need to be told just how these are to be interpreted in a given discussion.

I have not done a serious search of the use of the idea of typical or systematic ambiguity in mathematics and logic. The first employment of it that I've found (though without its being named in that way) was in Bertrand Russell's development of the theory of classes, relations and cardinal and ordinal numbers in his 1908 paper "Mathematical logic as based on the theory of types" [18]. How he used it there will be described in the next section. The idea is of course repeated and expanded in *Principia Mathematica*, beginning in Vol. I, from *20 on. But I did not find the words 'typical ambiguity' there either, except that in *65 there is talk of typically ambiguous symbols. Later uses by others are due, to begin with, to Quine ([16] and [17]) in connection with his system NF; the actual words 'typical ambiguity' are found in the second of these papers, pp. 132ff. The seminal work by Specker [19] aimed at proving the consistency of NF is entitled "Typical ambiguity"; that is tangentially related to what we are concerned with here in a way that will be briefly explained in the concluding section to this paper.

Russell used typical ambiguity in the theory of types to make sense of sentences like $Cls \in Cls$, which can be read as justifying talk in some sense of the class of all classes. In Section 3 this is generalized in a straightforward way to a treatment in simple type theory of statements of the form $A \in B$ where A and B are class expressions for which A is prima facie of the same or higher type than B. My main aim here is to extend that to mathematically interesting statements of the form $A \in B$ such as that the structure A we call the category of all categories is indeed a member of the class B of all categories. A few other test challenges of that sort are listed in Section 4. Then a version of typical ambiguity is formulated in an extension of ZF set theory in Section 5 and applied to the test cases in Section 6. The final Section 7 concludes with a discussion of the problem of justifying problematic membership statements when read literally.

[2]According to Godehard Link, "Shared opinion in the philosophy of language has it that in the case of indexicals like 'I', 'you', etc., it is not their meaning that varies with the context but rather their interpretation … This is considered to be different from lexical ambiguity ('bank'), syntactic ambiguity ('visiting relatives') or a combination of both ('I saw her duck')." [e-mail communication of 12 August 2003] I have received a similar comment from Thomas Wasow.

2. Russell's Uses of Typical Ambiguity

The theory of types described in [18] is the ramified theory, but what is essential for his use of typical ambiguity there is best explained in terms of the simple theory of types (STT) with the Axiom of Infinity. The types are thus indexed by the natural numbers $0, 1, 2, \ldots$ with the objects of type 0 interpreted as the individuals (infinite in number) and for each n, the objects of type $n+1$ interpreted as all classes of objects of type n. We use $x, y, z, \ldots$ as variables for any type $n \neq 0$ and a matching list of variables $X, Y, Z, \ldots$ of type $n+1$.[3] Only atomic formulas of the form $u \in W$ where u is of some type n and W of the next type $n+1$ are considered to be meaningful.

The first place that the issue of ambiguity comes up in [18] is on p. 251 (or p. 174 in its reprinting in [20]). For any type $n \neq 0$, he defines Cls[4] to be the class of all classes of objects of type $n-1$ and then writes:

> ... the proposition '$Cls \in Cls$' ... requires that 'Cls' should have a different meaning in the two places where it occurs. The symbol 'Cls' can only be used where it is unnecessary to know the type; it has an ambiguity which adjusts itself to circumstances.

A little further down he defines the empty class Λ and the universal class V and says that like Cls, these symbols are ambiguous "and only acquire a definite meaning when the type concerned is otherwise indicated." In our setting, these are particular objects of type n. Moving on from there Russell defines the Boolean operations $x \cup y$, $x \cap y$, and $-x$; this last is of course the complement relative to V. Later in the article he defines $Cl(x)$ to be the class of all subclasses of x; it is thus an object of type $n+1$. Extensionally, Cls is the same as $Cl(V)$.

The theory of types was created in order to save the logistic program from inconsistency, in particular to avoid Russell's paradox. The use of variables $x, y, z, \ldots$ of an indefinite type and the notions just explained illustrates the use, when necessary, of typical ambiguity to fall back on the underlying type structure for the security it provides ("having your cake") while regaining as much of the freedom of the informal theory of classes as possible ("eating it too").

Expanding the above form of STT to a theory of relations (as in Russell's theory) one can define the relation $x \sim y$ to hold when the classes x and y are in one-to-one correspondence. Then Russell defines $Nc(x)$ to be the class of all y such that $x \sim y$, i.e., $Nc(x)$ is the equivalence class of x under $\sim$; it is of type one higher than that of x. The class of cardinal numbers NC is then taken to be the class of all $Nc(x)$ for $x \in Cls$; it is of type $n+2$ for x of type n. The members $0, 1, 2, \ldots$ of NC are next introduced, respectively as $Nc(\Lambda)$, $Nc(\{0\})$, $Nc(\{0, 1\})$, and so on. Of these definitions, Russell writes:

[3] Russell uses $\alpha, \beta, \gamma, \ldots$ for class variables and $x, y, z, \ldots$ for variables for their elements. Our choice of $X, Y, Z, \ldots$ for class variables of type $n+1$ has been made so that one can more easily compare typical ambiguity in type theory with its use in set theory that we take up in Sec. 5 below.

[4] In [18], Russell wrote 'cls' where I write 'Cls'; the latter is used in *Principia Mathematica*.

> It has to be observed … that 0 and 1 and all the other cardinals … are ambiguous symbols, like *Cls*, and have as many meanings as there are types. To begin with 0: the meaning of 0 depends upon that of Λ, and the meaning of Λ is different according to the type of which it is the null class. Thus there are as many 0's as there are types; and the same applies to all the other cardinals.

Russell's partial way out of this embarrassing situation is to note that if classes x and y are of different types, for example x of a type n and y of type $n+1$, then we can speak of x and y having the same cardinal number or of one having a larger cardinal number than the other by comparing y with the class of singletons $\{u\}$ for $u \in x$. But this still does not get out of the fact that one has a multiplicity of representatives of the cardinals and in particular of the natural numbers. To an extent, the use of typical ambiguity is a way of saving face in this respect.[5]

The general problem that concerns us here is how to interpret expressions of the form $A \in B$ which are of indefinite type, but where the prima facie type of A is greater than or equal to that of B. Russell himself signaled this issue when suggesting how to deal with the pseudo statement $Cls \in Cls$; one simply interprets the second occurrence of '*Cls*' as being the class of all classes of type $n+1$, when the first occurrence is interpreted as the class of all classes of type n. A simpler statement which is meaningless on the strict account of type theory but which trades on the ambiguity of the symbols involved is $V \in V$, again legitimized by shifting the interpretation of the second 'V' one type higher.

There are other natural examples of this problem that Russell could have considered but did not. For example, the following statements seem reasonable (on the Axiom of Infinity):

$$\text{Inf} \in \text{Inf}, \quad \text{Fin} \in \text{Inf}, \quad \text{Inf} \notin \text{Fin}, \quad \text{Fin} \notin \text{Fin}, \tag{1}$$

where Fin, Inf are defined respectively to be the class of all finite classes and that of all infinite classes. So also are the statements:

$$\{1\} \in 1, \quad \{2\} \in 1, \quad \{1\} \notin 2, \quad \{1, 2\} \in 2. \tag{2}$$

[5] Link has suggested that something like the treatment of the meaning of indexicals in natural language (as quoted in footnote 3 above) could be applied to the notion of number in the theory of types: one could say that the meaning of the number terms is always the same as defined via an expression with bound class variables but that only the "typical" context (namely a given type level) determines which classes are being bound. While I am not advancing any theory of meaning here, in a sense this is what is done in Sec. 3, where the meaning is determined by a formula φ. But the additional problem comes when meeting such examples as (1), (2) in Sec. 2, so something more has to be done to specify the interpretations of the terms in formulas of the form $A \in B$, if both A and B contain the same terms, e.g., $\{1, 2\} \in 2$. This is what the disambiguation conventions are supposed to take care of.

3. Disambiguation of Membership Statements in Type Theory

Ambiguous expressions like Cls, Λ, V, 0, 1, ... can be assigned many types in STT; they are examples of expressions which are *stratified* in Quine's sense, i.e., result from erasing type indices from all the variables of an expression of STT (keeping variables of distinct types disjoint from each other). Given an expression S of STT, let $e(S)$ be the result of erasing all the type indices from variables of S. For A a stratified expression let type(A) be the least n for which there is an S of type n with $A = e(S)$; this is what we call the *prima facie type of* A. We shall confuse a stratified expression with the lowest expression of STT from which it results by erasing type distinctions. The disambiguation of expressions of the form $A \in B$ where the type of A is greater than or equal to that of B is first done here for the special case that they are of the same type and B is an expression of the form $\{x|\varphi(x)\}$, where φ is a stratified formula. Let φ^+ be stratified by replacing each variable in φ by the corresponding variable of next higher type, and then let B^+ be $\{X|\varphi^+(X)\}$. Then the *disambiguation convention* is simply:

(Dis 1) $A \in B$ means $A \in B^+$ when type(A) = type(B).

Thus for B of the form $\{x|\varphi(x)\}$, $A \in B$ is equivalent to $\varphi^+(A)$. For example, the statements (1) above may be inferred as an application of (Dis 1). Similarly, if A is of prima facie type one higher than $B = \{x|\varphi(x)\}$, we agree to the following disambiguation convention:

(Dis 2) $A \in B$ means $A \in B^{++}$ when type(A) = type(B) + 1.

Then, for example, the statements (2) above may be inferred as an application of (Dis 2).

The ambiguity in the kinds of formulas $A \in B$ considered here is typical because disambiguation is systematic using the principles (Dis 1) and (Dis 2) depending on the type of A relative to that of B, and similarly when A is of still higher type. It is also typical because it doesn't really depend on the choice of type assignments on the basis of which the type of A is measured in comparison with that of B, as long as all types are shifted by the same amount. This is because of the straightforward:

Theorem 1. *If a sentence θ is provable in* STT *then so also is θ^+.*

It follows that if $A \in B$ is provable in STT where A, B are given by closed terms, then so also is $A^+ \in B^+$. More importantly, if $A \in B$ is provable then every property of elements of B also holds of A. This is formulated as:

Theorem 2 (Transfer rule). *For closed terms A, B, C with $B = \{x|\varphi(x)\}$ and $C = \{x|\psi(x)\}$ and type(A) = type(B), if $A \in B$ and $B \subseteq C$ are provable in* STT *then so also is $A \in C$.*

Proof. This is because $A \in B$ means $\varphi^+(A)$. Since $\forall x[\varphi(x) \to \psi(x)]$ is a theorem, so also is $\forall X[\varphi^+(X) \to \psi^+(X)]$, so $\psi^+(A)$ holds, i.e., $A \in C$ holds by the disambiguation convention. □

4. Some Mathematical Challenges

All the preceding is quite obvious; moreover the applications are of limited mathematical interest, because STT doesn't lend itself to the flexible expression of mathematical properties in practice. Here are some statements of the form $A \in B$ with the prima facie type of A greater than or equal to that of B that are intuitively true in a naive theory of structures but cannot be verified directly in current systems of type theory or set theory. The aim is to make sense of them by some form of typical ambiguity.

Note that in the following we use $(.\,,.)$ for the pairing operation, which is iterated to form n-tuples. We assume understood the mathematical notions involved, cf. [13] for the examples 4.2–4.4.

4.1 Let P be the class all partially ordered structures and let S be the substructure relation. Then $(P, S) \in P$.

4.2 Let Set be the category of all sets, AbGrp the category of all Abelian groups, Top the category of all topological spaces, etc., and let CAT be the class of all categories. Then, as should be, Set $\in$ CAT, AbGrp $\in$ CAT, Top $\in$ CAT, etc.

4.3 Also Cat $\in$ CAT where Cat $=$ (CAT, FUNCT, $\circ$) is the *category of all categories*, whose objects are all categories and whose morphisms are the functors between categories, and $\circ$ is the partial operation of composition of functors.

4.4 If $A \in$ CAT and $B \in$ CAT then $B^A \in$ CAT where

$$B^A = (\mathrm{FUNCT}(A, B), \mathrm{NAT}(A, B), \circ),$$

whose objects are the class FUNCT(A, B) of all functors from A to B and whose morphisms are the natural transformations between functors and where $\circ$ is the operation of composition of such transformations.

4.5 Let BA be the class of all Boolean algebras. Then $(\wp(V), \cup, \cap, -, \varnothing, V) \in$ BA where V is the universal class and $\wp(V)$ is the class of all subclasses of V.[6]

[6] This example was suggested to me by Godehard Link.

5. Typical Ambiguity in a System of Set Theory with Universes

Let L be the language of ZF set theory, using variables $x,\ y,\ z,\ \ldots$ for sets. Adjoin to L constants U_n for $n = 0,\ 1,\ 2,\ \ldots$ for an increasing sequence of *reflective universes*; the resulting language is denoted $\mathrm{L}(U_{<\omega})$, where we use $U_{<\omega}$ to indicate the sequence of U_n for $n < \omega$. Each U_n is supposed to be a set which is reflective in the sense that—speaking model-theoretically—it forms an elementary substructure of $(V, \in)$ when the membership relation is restricted to U_n. As usual, to express this, we use the operation φ^a of forming the relativization of all quantifiers in the formula φ of L to a; we also write $\mathrm{Rel}(\varphi, a)$ for the resulting formula. The universes are also supposed to be *supertransitive*, i.e., transitive and closed under the power set operation $\wp$; for the latter it is sufficient to assume that universes are closed under subsets of members. The system of ZF, resp. ZFC, with universes satisfying these properties is denoted $\mathrm{ZF}/U_{<\omega}$, resp. $\mathrm{ZFC}/U_{<\omega}$. More officially, the axioms are as follows.

Axioms of $\mathrm{ZF}/U_{<\omega}$

I. All the axioms of ZF in L, and for each $n = 0, 1, 2, \ldots$:

II. $U_n \in U_{n+1}$.

III. $\mathrm{Trans}(U_n)$.

IV. $\forall x \forall y[y \in U_n \wedge x \subseteq y \rightarrow x \in U_n]$.

V. For each L formula $\varphi(x_1, \ldots, x_k)$,
$\forall x_1 \ldots \forall x_k\{x_1, \ldots, x_k \in U_n \rightarrow [\mathrm{Rel}(\varphi, U_n)(x_1, \ldots, x_k) \leftrightarrow \varphi(x_1, \ldots, x_k)]\}$.

The Axioms of $\mathrm{ZFC}/U_{<\omega}$ are obtained by adding the Axiom of Choice, AC.

Theorem 3. *$\mathrm{ZF}/U_{<\omega}$ is a conservative extension of* ZF.

Proof. This is by a straightforward extension of the result of Montague and Vaught [14] which in turn modifies Levy's well known reflection principle argument, see also [5]: Sec. 2. We give here a brief sketch of the ideas involved. Note that for conservation we need only show how, given any finite set of axioms of $\mathrm{ZF}/U_{<\omega}$ involving only symbols U_i for $i = 0, \ldots, n$ for some n, to define sets in ZF satisfying the given axioms for these universes. Working informally in ZF, let V_α be the αth set in the cumulative hierarchy, i.e., for each α, V_α is the union of $\wp(V_\beta)$ for all $\beta < \alpha$. For any set x, the rank of x, $\rho(x)$, is the unique α such that $x \in V_{\alpha+1} - V_\alpha$. One associates with each existential formula $(\exists x)\psi(x, y_1, \ldots, y_m)$ of L an m-ary function F_ψ whose value $F_\psi(y_1, \ldots, y_m)$ for each $y_1, \ldots, y_m$ is the set of all x of least rank such that $\psi(x, y_1, \ldots, y_m)$ holds. Thus

$$(\exists x)\psi(x, y_1, \ldots, y_m) \rightarrow (\exists x \in F_\psi(y_1, \ldots, y_m))\psi(x, y_1, \ldots, y_m).$$

The F_ψ act like Skolem functions, but without needing the Axiom of Choice to pick a specific value of witness for the existential quantifier. By the *rank-hull* of a set b we mean the least V_α such that $b \subseteq V_\alpha$. For any finite set Q of formulas $(\exists x)\psi(x, y_1, \ldots, y_m)$, by the *Skolem-rank-hull* of a set b relative to Q, in symbols $H_Q(b)$, we mean the least V_α such that $b \subseteq V_\alpha$ and such that for each $(\exists x)\psi(x, y_1, \ldots, y_m)$ in Q and each $y_1, \ldots, y_m \in V_\alpha$, $F_\psi(y_1, \ldots, y_m) \subseteq V_\alpha$. $H_Q(b)$ is obtained as the union of bj for $j < \omega$, where $b_0 = b$ and each b_{j+1} is the rank hull of [b_j union all $F_\psi(y_1, \ldots, y_m)$ for all formulas $(\exists x)\psi(x, y_1, \ldots, y_m)$ in Q and all $y_1, \ldots, y_m \in b_j$]. For simplicity, assume the formulas of L are generated from atomic formulas $x \in y$ and $x = y$ by $\neg$, $\wedge$ and $\exists$. Now, given the finite set of axioms of ZF/$U_{<\omega}$ to be modeled in ZF, let S be the closure under subformulas of all the L formulas in the given set and let Q consist of all existentially quantified formulas $(\exists x)\psi$ in S. Then it is proved by formula induction that for each set b and each φ in S,

$$\forall x_1 \ldots \forall x_k \{x_1 \ldots x_k \in H_Q(b) \rightarrow [\mathrm{Rel}(\varphi, H_Q(b))(x_1 \ldots x_k) \leftrightarrow \varphi(x_1 \ldots x_k)]\}.$$

Since also each $H_Q(b)$ is supertransitive by construction, we can define U_i for $i = 0, \ldots, n$ by taking $U_0 = H_Q(0)$ and for each $i < n$, $U_{i+1} = H_Q(U_i \cup \{U_i\})$. □

NB. If θ is any sentence of L then adding θ as an axiom maintains conservativity by this theorem.

Corollary. ZFC/$U_{<\omega}$ *is a conservative extension of* ZFC.

Note that each reflective universe U_n satisfies all the axioms of ZF by taking closed φ in Axiom V to be any one of these axioms. It follows by supertransitivity that each universe contains the empty set 0 and the set ω of finite ordinals, is closed under unordered pair, union and power set; moreover, these are absolute, i.e., have the same values in the universe as in the universe of all sets. Moreover, each universe U_n is closed under the Separation Axiom and Replacement Axiom schemes. The latter can be thought of as follows: if f is any function in the set-theoretic sense of the word that is *defined in* L (with respect to any given parameters in U_n) and if $a \in U_n$ and $f : a \rightarrow U_n$, then $\{f(x) : x \in a\} \in U_n$. If one wishes to drop here the assumption that f is defined, allowing it to be *any function*, then $\rho(U_n)$ would have to be a strongly inaccessible cardinal, and assumption of that would no longer hold conservatively over ZF; instead we would have to strengthen ZF by the assumption of the existence of infinitely many strongly inaccessible cardinals. This might be needed for some applications (see the next section), but is not assumed here.

We now consider to what extent application of typical ambiguity in the system ZF/$U_{<\omega}$ (or the same augmented by AC as dictated by specific needs) can be used to meet the mathematical challenges of the preceding Sec. 4.[7] Looked at informally,

[7] Something like this was suggested as follows by Kurt Gödel in a letter to Paul Bernays dated 1 January 1963: "I find it interesting that you speak …of the 'newer abstract disciplines of mathematics' as something lying outside of set theory. I conjecture that you are thereby alluding to the concept of category and to

what we have to deal with to begin with in the examples 4.1–4.3 are membership relations of the form $A \in B$, where B is a class and A is a (possibly) many-sorted relational structure each of whose domains and relations are classes. In all the cases considered, B is given as $\{x|\varphi(x)\}$ for some L formula φ; the classes which are the constituents of A are also defined in L. We shall relativize the concepts involved in A and B to universes. By a *typical reflective universe* U, we mean *any* U_n; then by the *next universe* U^+ we mean U_{n+1}. The ***first step*** in interpreting the problematic membership statements relative to any such universe U is to *identify classes with the corresponding subsets of* U. The ***second step*** is to *identify sets with the members of* U. Note that since the classes making up the structure A are subsets of U, we have $A \in U^+$. Now, to make sense of $A \in B$, the ***third step*** is to *re-identify* B *with the class of all sets in* U^+ *that satisfy the definition of* B; call this B^+. More precisely, if B is given formally as $\{x|\varphi(x)\}$, its first identification is with $\{x \in U|\varphi^U(x)\}$ which is the same as $\{x \in U|\varphi(x)\}$; then, the re-interpretation B^+ of B is simply defined to be $\{x \in U^+|\varphi(x)\}$. This leads to the disambiguation convention:

(Dis 3) $\qquad A \in B$ means $A \in B^+$.

To take the simplest example, 4.1, P is defined to be the class of structures (x, y) where $y \subseteq x^2$ is a partial ordering of x, and S is defined to be the substructure relation $(x, y) \subseteq (z, w)$ on P which holds just in case $x \subseteq z$ and $y = w \cap x^2$. With the variables x, y, z, w taken to range over the sets in any typical reflective universe U, we have that P and S are subsets of U. According to the disambiguation principle (Dis 3), $(P, S) \in P$ means that $(P, S) \in P^+$, where P^+ consists of all those sets (X, Y) with X, Y in U^+ such that $Y \subseteq X^2$ and Y is a partial ordering of X. It is readily verified that in this case, we do indeed have $(P, S) \in P^+$, so it makes sense by the disambiguation principle to assert that $(P, S) \in P$. In words: the class of all partially ordered structures (in a typical universe) together with the substructure relation (in that universe) forms a partially ordered structure (in the next universe).

There is more to be said about the above disambiguation principle, again illustrated by reference to 4.1. By the reflection axiom V of $\mathrm{ZF}/U_{<\omega}$, P and P^+ share exactly the same properties expressed in the language L of set theory, and *each* relativized to any typical reflective universe can be considered as a *surrogate* of the class of all sets (x, y) which form partial ordering structures. Moreover, consider any property $\Psi(x, y)$ formulated in L which is proved to be true of all such (x, y). Let C be defined as the class of all (x, y) such that $\Psi(x, y)$ holds. This definition is taken to be ambiguous, i.e., it can be thought of as pertaining to the class of all sets, or relativized to any universe. Our assumption is that $P \subseteq C$ has been proved in ZF, i.e., when we relativize to the class of all sets. It follows that $P^+ \subseteq C^+$ when P is interpreted as the

the self-applicability of categories. But it seems to me that all of this is contained within a set theory with a finitely iterated notion of class, where reflexivity results automatically through a 'typical ambiguity' of statements." (The source for the German original of this letter is in the Bernays archives at the ETH in Zürich; the full letter is to be found, along with much else between Gödel and Bernays, in the collection of correspondence in [9].)

class of all partially ordered structures in any typical reflective universe U. From the disambiguation principle, it follows that $(P, S) \in C$. In other words, (P, S) *shares all the properties that are verified of all partially ordered structures*; this is an example of the *transfer rule* in this framework. More generally, we have:

Theorem 4 (Transfer rule). *If $A \in B$ and $B \subseteq C$ are provable in* ZF *where B, C are class abstracts then $A \in C$ is provable there too.*

Proof. For B given formally as $\{x|\varphi(x)\}$ and C as $\{x|\psi(x)\}$, $B \subseteq C$ is read as $\forall x[\varphi(x) \rightarrow \psi(x)]$, and what it means for $B \subseteq C$ to be provable in ZF is that $\forall x[\varphi(x) \rightarrow \psi(x)]$ is provable in ZF. Then by the reflection axiom V, given any universe U, $(\forall x \in U^+)[\mathrm{Rel}(\varphi, U^+)(x) \rightarrow \mathrm{Rel}(\psi, U^+)(x)]$, i.e., $B^+ \subseteq C^+$. From provability of $A \in B$, which means that $A \in B^+$ by (Dis 3), it follows that $A \in C^+$, i.e., $A \in C$, again by the disambiguation convention. □

To look at things more generally, consider a type structure over sets, where sets are considered to be at type level 0, classes of sets at type level 1, classes of classes of sets of type level 2, and so on. We regard the type level of a structure A to be the same as the maximum type level of the domains, relations, operations and individuals that make up that structure. Thus in the examples 4.1–4.3 of problematic membership statements of the form $A \in B$ the type level of both A and B is 1. But in the example 4.5, the structure $A = (\wp(V), \cup, \cap, -, \oslash, V)$ is of type level 2 since the class V of all sets is of type level 1 and $\wp(V)$, the class of all subclasses of V, is of type level 2. On the other hand the type level of B—in this case the class BA of all sets which are Boolean algebras—is still 1. But unlike the case of type theory, where to disambiguate $A \in B$ we had to go to $A \in B^{++}$ when the type level of A is one higher than that of B, we can still disambiguate by (Dis 3). For, when we interpret the sets as ranging over any typical universe U, A is interpreted as being the structure $(\wp(U), \cup, \cap, -, \oslash, U)$ and since U and $\wp(U)$ belong to U^+, the structure A in this case also belongs to U^+, and in fact is a member of the subset BA^+ of U^+ consisting of all structures in U^+ which satisfy the conditions to be a Boolean algebra. Thus it makes sense by (Dis 3) to say that $A \in B$ holds in this case. The advantage over ordinary type theory given to us by the set-theoretical framework is that the type level of U^+ over U measured in terms of the cumulative hierarchy is infinite. A precise explanation of how (Dis 3) may be applied more generally would require the introduction of an extension L^+ of L by variables of classes, which are assumed (unlike Gödel–Bernays or Morse–Kelley theories) to satisfy all the axioms of ZF. Then for any application, the ZF objects are interpreted as ranging over a universe U and the classes are interpreted as ranging over the next universe U^+. For the purposes here, it is sufficient to see how (Dis 3) works in a few specific cases, like that of 4.5 just discussed, or 4.4 to be treated in the next section.

6. Category Theory in ZF/$U_{<\omega}$

The reader is assumed to be familiar with the basic notions of category theory. A standard reference is [13]. We can take categories to be structures of the form $A = (O, M, C)$ where O is the class of objects of A, M its class of morphisms and C the composition of morphisms in A; C is a partial operation on M^2 to M considered as a three-placed relation on M. Actually, we can identify O with the class of identity morphisms, suitably defined, and the domain and codomain of a morphism with its left and right identities, so it is sufficient to consider structures of the form (M, C). Alternatively, and more intuitively, we can take categories to be structures of the form $A = (O, M, C, D_0, D_1, I)$ where C is as before, D_0 and D_1 are maps from M to O giving the domain and codomain respectively, and I is a map from O to M. As usual we write $f : x \to y$ or $x \to^f y$ in A when $f \in M$, $D_0(f) = x$ and $D_1(f) = y$; id_x for $I(x)$; and $fg = h$ when $(f, g, h) \in C$. When it is necessary to compare objects from one category with another we subscript the terms of A by 'A' as $A = (O_A, M_A, C_A, D_{0A}, D_{1A}, I_A)$. A is a *large category* if O_A and M_A are both classes. A is said to be *locally small* if for any objects x, y of A the class $H_A(x, y) = \{f | f \in M_A \wedge f : x \to y\}$ is a set; it is *small* if A itself is a set.

The following are standard examples of locally small categories.

The category of all sets. Set is the category whose objects are just the sets and whose morphisms are all triples $f = (u, x, y)$ where u is a function (in the usual set-theoretical sense) whose domain is x and range is contained in y. Then $D_0(u, x, y) = x$ and $D_1(u, x, y) = y$. The composition fg is defined for $f = (u, x, y)$ and $g = (v, z, w)$ just in case $y = z$, in which case $fg = (u; v, x, w)$ where $u; v$ is the relational composition of u and v. Finally $I(x)$ is taken to be (u, x, x) for each x, where u is the identity function from x to x.

The category of Abelian groups. AbGrp is the category whose objects are the sets which are Abelian groups and whose morphisms are group homomorphisms with specified domain and codomain. This is then spelled out as in 6.1.

The category of all topological spaces. Top is the category whose objects are the members of U that are topological spaces, treated similarly.

The category of all categories. Cat is the category whose objects are the sets (o, m, c, d_0, d_1, i) that satisfy the conditions to be a category, and whose morphisms are the functors between any two such categories. Cat itself has the form (CAT, FUNCT, ...) where $\mathrm{CAT} = O_{\mathrm{Cat}}$ and $\mathrm{FUNCT} = M_{\mathrm{Cat}}$.

Now by (Dis 3) we can make sense of the statements in 4.2 and 4.3 that Set $\in$ CAT, AbGrp $\in$ CAT, Top $\in$ Cat and even Cat $\in$ CAT, i.e., (CAT, FUNCT, ...) $\in$ CAT. Disambiguated, these statements mean that relative to any typical reflective universe U, each of Set, AbGrp, Top, and Cat is an element of CAT^+, the class of all sets $A = (O, M, \ldots)$ in the next universe U^+ which satisfy the conditions to be a category. Furthermore, the transfer theorem assures us that *any properties that apply to all*

members of Cat*, i.e., the class of all* ***small*** *categories, also hold for the* ***large*** *categories* Set*,* AbGrp*,* Top*,* Cat*, etc.*

The category of all functors between two given categories. When we turn to making sense of the statement 4.4 that $B^A \in \mathrm{CAT}$ where A and B are large categories we meet the problem that the prima-facie type level of B^A is higher than that of both A and B. In Mac Lane's terminology, B^A is beyond large, sometimes called *superlarge*. But from the point of view of the method of disambiguation proposed in the preceding section, the issue here is no different from that met with the example 4.5 concerning the Boolean algebra on $\wp(V)$. But there are mathematical aspects of example 4.4 that are worth a closer look. Consider the usual set-theoretical definition of B^A: its objects are $\mathrm{FUNCT}(A, B)$, i.e., the class of all functors $F : A \to B$, and its morphisms are natural transformations $\eta : F \to G$ between such functors. We can take functors to be classes which are functions in the usual set-theoretical sense, so for each $x \in O_A$, $F(x) \in O_B$; as usual, to be a functor, F is required to preserve composition and identity morphisms, i.e., for $f, g \in M_A$ and $x \in O_A$, $F(fg) = F(f)F(g)$ in B and $F(i_x) = i_{F(x)}$. By a natural transformation $\eta : F \to G$ between two such functors $F : A \to B$ and $G : A \to B$ is meant a map from O_A to M_B such that the following diagram is commutative for any $x, y \in O_A$, whenever $x \to^u y$ in A:

$$\begin{array}{ccc} F(x) & \xrightarrow{F(u)} & F(y) \\ {\scriptstyle \eta(x)}\downarrow & & \downarrow{\scriptstyle \eta(y)} \\ G(x) & \xrightarrow{G(u)} & G(y) \end{array}$$

Composition of natural transformations is defined in an obvious way. Though, as noted above, B^A is no longer a large category when A and B are large, we can still make use of the disambiguation convention (Dis 3) as at the end of the preceding section to make sense of the statement that it is a member of CAT, since, according to it, this simply means that relative to any typical universe U, $B^A \in \mathrm{CAT}^+$. [8] There is no problem with this, since now with the O_A, M_A, O_B, and M_B regarded as subsets of a given universe U, the sets O_C and M_C belong by the above definition to U^+ for $C = B^A$. Again, the transfer theorem assures us *that any property of all* ***small*** *categories also applies to the* ***superlarge*** *category* B^A.

7. Discussion: Axiomatic Foundations of Category Theory

One of the usual foundations of category theory, due to Mac Lane [11], takes as its setting the language of the Bernays–Gödel theory of sets and classes; this allows us to talk about two kinds of categories, those that are *small*, i.e., are sets, and those that are

[8] This is in accord with Gödel's idea quoted in footnote 7.

large, i.e., are classes which are not sets. In such a foundation, when A and B are both large categories, such as $A = \text{AbGrp}$ and $B = \text{Set}$, there is no place to locate B^A, since it is now of a type higher than the classes in the sense of the language of BG. Thus one often is forced to restrict this construction to the case that A is small, so that the functors from A to B are all sets, and B^A is at most a large category in the BG sense. The preceding shows that that problem is not met in the $\text{ZF}/U_{<\omega}$ setting presented here. Thus, for example, we can state the Yoneda Lemma as a natural equivalence between any locally bounded (i.e., locally small) category A and the category $\text{Set}^{A'}$, where A' is the opposite category to A.

Now it is to be noticed that we never needed more than two universes U and U^+ to take care of disambiguation in the various examples 4.1–4.5. Thus, we could just as well have restricted the theory to the part of $\text{ZF}/U_{<\omega}$ that concerns only U_0 and U_1. One could go even farther, by dealing with *only one reflective universe* U, of which (axiomatically) the full universe V is taken to be an elementary extension, thus with U being treated like U_0 and V like U_1; call this form of the theory ZF/U_0.[9] Large categories have as surrogates subcategories of U_0, and functor categories for them simply sit as sets in V. On the face of it, this would seem to make it similar to the suggestion of Mac Lane [12] according to which one universe suffices. But there is an essential difference here in the further use of the reflection Axiom V to insure that there is nothing special about the choice of the universe U_0; with respect to properties formulated in set-theoretical terms, it is indistinguishable from the full universe. Thus any property established for subcategories of U_0 holds of all categories, in particular the category Cat and the functor categories of 4.4.

Actually, the idea of using a theory like ZF/U_0 or $\text{ZFC}/U_{<\omega}$ as a foundational framework for category theory is an old one that I first elaborated in the paper "Set-theoretical foundations of category theory" [5].[10] What $\text{ZF}/U_{<\omega}$, resp. $\text{ZFC}/U_{<\omega}$, provides as an advantage is the ability to explain more directly the disambiguation relative to any typical reflective universe in terms of the next universe. The adequacy of ZFC/U_0 as a framework for working category theory was considered with respect to some prominent test cases, including: (i) Yoneda Lemma, (ii) Freyd Adjoint Functor Theorem, and (iii) the functors Ext_n in homological algebra. These, and others, should be re-considered as test cases for a more flexible development in the system $\text{ZFC}/U_{<\omega}$.

In some cases we may need somewhat stronger hypotheses; for example (as explained in [5]), it appears that the Kan Extension theorem requires the rank of the typical universe U being considered to be a strongly inaccessible cardinal. To deal with such cases we shall need to add as an assumption to our base system of set theory

[9] That was the form in which typical ambiguity in set theory was presented in a draft of this paper. The reasons for the shift to a theory with many reflective universes are given below.

[10] A similar proposal has more recently been made by F. A. Muller [15], taking an extension of the theory of sets and classes due to Ackermann, as the basic framework; this system is known to be interpretable in ZF/U_0 by interpreting the sets to be the members of U_0 and the classes to be the sets in V. Muller incorrectly asserts that my paper [5] required the assumption of inaccessible cardinals. For a critical review by Andreas Blass of Muller (2001) and elaboration of its relation to my earlier work, see http://www.ams.org/mathscinet/review 2002k:03008.

that beyond any ordinal there is a strongly inaccessible cardinal. (Such an assumption is considered to be innocuous by working set theorists.) A proposed foundation for category theory ascribed to Grothendieck makes use of this assumption; by a universe on that approach is meant any V_α for α a strongly inaccessible cardinal. What that foundation does not explain is why such universes may be considered typical. That is what the reflection Axiom V adds to the Grothendieck approach. But it is useful to see when the much weaker system ZFC/$U_{<\omega}$, without the assumption of any inaccessibles, suffices for various parts of the development of category theory; in fact, that assumption seems to be rarely needed.

Another, general, reason for dealing here with the system with a sequence of reflective universes rather than a single one has been to show how the handling of typical ambiguity in set theory parallels, in part, that in the theory of types as dealt with above in Sec. 3.

There are other candidates for foundational frameworks for category theory employing typical ambiguity which would seem to have some advantages over ZF/$U_{<\omega}$ with or without AC, especially with respect to the treatment of functions. One is a form of operational set theory, in which the system is extended further by variables for a partial combinatory algebra over the universe.[11] Another is to use one of the systems of Explicit Mathematics introduced in [7] and studied in a number of publications since then. There the operational structure in the form of a partial combinatory algebra on the universe of all individuals is taken as basic. For a system of that kind to work we would have to extend the formalism by symbols for typical reflective universes with the appropriate axioms. Both treatments allow B^A to be interpreted directly as a class of type level 1 rather than type level 2. I have done some experimentation with both of these approaches, but have not yet brought the work to a definitive form.

8. Trying to Have Your Cake and Eat It Too: Naive Category Theory

Minimally, what one is after is to have a demonstrably consistent foundational framework T for mathematically interesting cases of self-membership such as provided by category theory. But, more than consistency, one would want (as with ZF/$U_{<\omega}$) conservativity over an accepted framework (in that case over ZF); all that is done here, though not without some clumsiness in the applications. Ideally, what one is after—and that is *not* done here—is to provide a framework meeting these criteria in which such objects as CAT exist in the universe of discourse and which are such that (CAT, ...) $\in$ CAT is *literally true*, and *not reinterpreted* according to context.

[11] See http://math.stanford.edu/ Feferman/papers/OperationalST-I.pdf for a draft article featuring such a system. (NB. Theorem 4(i) p. 5 needs correction.) Incidentally, Vidhyanath Rao has informed me that there are some theorems which require global choice for their proof. The system of operational set theory referred to here incorporates a global choice operator.

For a discussion of the desiderata for that kind of foundation, see my paper "Categorical foundations and foundations of category theory" [8]. We have yet to obtain a satisfactory solution of the problem posed by naive category theory in that form.

At first sight, one direction in which such a solution might be sought is to make use of theories like that of Aczel [1] as exposited in [4], in which the Foundation Axiom is replaced by the Anti-Foundation Axiom AFA. That certainly allows many examples of self membership such as $a \in a$ and $(a, b) \in a$, and so on, with interesting applications when suitably elaborated. In Ch. 20 of that book there is a proposed extension SEC_0 of the theory of sets in which class variables are adjoined, that also allows some cases of self-membership between classes such as $A \in A$ and $(A, B) \in A$; *grosso modo*, such statements are needed to deal with the kinds of mathematical applications from category theory considered here. Some of the above desiderata are met by that system, namely that of consistency relative to ZFC, of which it is an extension. But, as the authors themselves point out, the mathematical usefulness of SEC_0 in general remains to be established.[12]

Another direction in which such a solution might be sought is via some form of stratified theories like Quine's NF. Though that is still not known to be consistent, the system NFU with urelements allowed was shown to be consistent by Jensen [10], and he also showed how it can be beefed up to include, conservatively, ZFC. However, the formalism of NFU is not as it stands suitable to define $\{x | x$ is a relational structure of signature $\sigma\}$ where σ is any specified signature. For example, if x is to be of the form (y, z) with $z \subseteq y^2$ we need to assign to the elements of z, which are ordered pairs, the same type level as the elements of y, and thus ordered pairs of elements of a set need to be assigned the same type level as the elements of that set. None of the usual definitions of ordered pair in NFU works to do that. However, one can formulate a simple extension NFUP of NFU in which pairing is taken as a basic operation, and stratification is modified so as to allow us to assign the same type to an ordered pair as to its terms. By an adaptation of Jensen's proof, one can establish the consistency of NFUP and again one can beef it up to obtain a version of it conservative over ZFC. I carried this out in an unpublished MS, "Some formal systems for the unlimited theory of structures and categories" (abstract in [6]).[13] This system indeed serves to literally verify examples like 4.1–4.4. But it has other defects as a proposed foundation of naive category theory. One is that there is no way that one can establish existence of the cartesian product $\prod x_i (i \in I)$ of a collection $\{x_i | i \in I\}$ of sets, since the collection must be given by a function g with $g(i) = x_i$ for each $i \in I$, and the elements of the cartesian product f must be of the form $(i, f(i))$ with $f(i) \in g(i)$. Thus there is no stratified type assignment for pairing which allows one to deal with both f and g simultaneously. On the other hand, there is no obvious way to obtain a consistent

[12] According to one of the referees, the work of Barwise and Moss on SEC_0 has been continued by Alexandru Baltag in his doctoral thesis, and one publication resulting from that is [2]. I do not know how relevant it may be to the questions dealt with here.

[13] The proof makes use of the existence of two inaccessible cardinals, an assumption which, as remarked above, is regarded as innocuous by working set-theorists.

extension of NFU allowing stratification of pairs where the terms of a pair are of prima facie mixed type.

As the title of Specker [19] attests, there is of course a non-trivial connection of the consistency problem for stratified systems with typical ambiguity. Specker considered one form of typical ambiguity in STT to be given by the scheme $(\varphi \leftrightarrow \varphi^+)$ for all sentences φ of type theory; he showed that NF is consistent just in case STT is consistent when augmented by this scheme. One way to insure that would be to seek models of type theory in which there is a shifting endomorphism which takes each type level to its successor level, or a model of STT allowing negative types in which there is a type shifting automorphism. Jensen succeeded in using Specker's idea in his proof of consistency of NFU, by combining it with the Ehrenfeucht–Mostowski theorem on the existence of models with many automorphisms. But even if the consistency of NF were established via Specker's theorem or by some alternative approach, NF would not, as it stands, provide the interpretation of structural notions needed to insure literal self-membership in the sense of such examples as 4.1–4.4, as well as to satisfy the additional criteria of [8] for ordinary mathematical constructions such as that for product above.

Acknowledgement. Invited lecture for the conference, One Hundred Years of Russell's Paradox, University of Munich, June 2–5, 2001. I wish to thank Godehard Link for his exceptional work in organizing this conference and for his personal assistance in connection with it. I have received useful comments on this paper from Tim Fernando, Godehard Link, Karl-Georg Niebergall and two anonymous referees.

References

[1] Aczel, Peter: 1988. *Non-Well-Founded Sets*. Stanford: CSLI Publications no. 14.

[2] Baltag, Alexandru: 1998. STS: A structural theory of sets. *Advances in Modal Logic* 2: 1–34.

[3] Bartlett, John: 1980. *Familiar Quotations*, Fifteenth edition. Edited by E. M. Beck. Boston: Little, Brown and Co.

[4] Barwise, Jon and Lawrence Moss: 1996. *Vicious Circles*. Stanford: CSLI Publications no. 60.

[5] Feferman, Solomon: 1969. Set-theoretical foundations of category theory (with an appendix by G. Kreisel). In: M. Barr *et al*. (eds.), *Reports of the Midwest Category Seminar III*, Lecture Notes in Mathematics 106, 201–247.

[6] Feferman, Solomon: 1974. Some formal systems for the unlimited theory of structures and categories (abstract). *J. Symbolic Logic* 39: 374–375.

[7] Feferman, Solomon: 1975. A language and axioms for explicit mathematics. In: J. Crossley (ed.), *Algebra and Logic*, Lecture Notes in Mathematics 450, 87–139.

[8] Feferman, Solomon: 1977. Categorical foundations and foundations of category theory. In: R. E. Butts and J. Hintikka (eds.), *Logic, Foundations of Mathematics and Computability Theory*, vol. 1, Dordrecht: Reidel, 149–169.

[9] Gödel, Kurt: 2003. *Collected Works, Volume IV. Correspondence A–G*. Edited by S. Feferman *et al*., Oxford: Oxford Univ. Press.

[10] Jensen, Ronald: 1969. On the consistency of a slight (?) modification of Quine's *New Foundations*. In: D. Davidson and J. Hintikka (eds.), *Words and Objections: Essays on the Work of W. V. O. Quine*, Dordrecht: Reidel, 278–291.

[11] Mac Lane, Saunders: 1961. Locally small categories and the foundations of mathematics. In: *Infinitistic Methods*, Oxford: Pergamon Press, 25–43.

[12] Mac Lane, Saunders: 1969. One universe as a foundation for category theory. In: M. Barr *et al*. (eds.), *Reports of the Midwest Category Seminar III*, Lecture Notes in Mathematics 106: 192–200.

[13] Mac Lane, Saunders: 1971. *Categories for the Working Mathematician*. Berlin: Springer.

[14] Montague, Richard and Robert L. Vaught: 1959. Natural models of set theories. *Fundamenta Mathematicae* 47: 219–242.

[15] Muller, Frederik A.: 2001. Sets, classes, and categories. *British J. Philosophy of Science* 52: 539–573.

[16] Quine, Willard V.: 1937. New foundations for mathematical logic. *American Mathematical Monthly* 44: 70–80.

[17] Quine, Willard V.: 1938. On the theory of types. *J. of Symbolic Logic* 3: 125–139.

[18] Russell, Bertrand: 1908. Mathematical logic as based on the theory of types. *American Mathematical Monthly* 30: 222–262. Reprinted in [20]: 150–182.

[19] Specker, Ernst: 1962. Typical ambiguity. In: E. Nagel *et al*. (eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, Stanford: Stanford Univ. Press, 116–124.

[20] van Heijenoort, Jean (ed.): 1967. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard Univ. Press.

[21] Whitehead, Alfred N. and Bertrand Russell: 1925. *Principia Mathematica*, vol. I, second edition. Cambridge: Cambridge Univ. Press.

Department of Mathematics
Stanford University
Stanford, CA 94305-2125
USA

E-mail: sf@csli.stanford.edu

Is ZF Finitistically Reducible?

Karl-Georg Niebergall

Abstract. It follows from work done by J. Mycielski that each consistent r.e. first-order theory T is proof-theoretically reducible to a locally finite theory T'. I present a different method for obtaining the same result. A consequence of these results is: *If each locally finite theory is finitary, then ZF is finitistically reducible*. Since the premiss "Each locally finite theory is finitary" may be accepted by some proof theorists, we have the problem that ZF would turn out to be finitistically reducible.

1. Introduction

I think that the answer to the question posed in the title of this text should be "No". Since the expression "finitistically" echoes the word "finite" from ordinary language, this should be intuitively obvious. Even without having precise definitions of "T is finitistically reducible" and "S is a finitistic theory"[1] at hand—

$$S \text{ is finitary} \iff S \text{ does not make assumptions of infinity}$$

seems *prima facie* plausible from this perspective. Assumptions of infinity are made by ZF, however, to such a high degree that it would furthermore stretch our understanding of "reducible" if ZF were declared to be finitistically reducible. It is reassuring that in investigations belonging to proof theory the same judgement can be found: "For example (to be extreme again), the system ZF, which is justified by the uncountable infinitary framework of Cantorian set theory, is not reducible to any finitarily justified system." ([6]: 7).

At the same time, J. Mycielski (in [29] and [30]) claims that ZF is *syntactically isomorphic* to a *locally finite* theory FIN(ZF), which, in Mycielski's words, "eliminates ideal (infinite) objects from the proofs of properties of concrete (finite) objects".[2] And S. Lavine, in proposing an approach to the foundations of mathematics (see [24] and

[1] I use "S is a finitistic theory" and "S is a finitary theory" as synonyms.

[2] "It is the purpose of this paper to construct in a uniform way for any consistent theory T a locally finite theory FIN(T) which is syntactically (in a sense) isomorphic to T." ([30]: 59). And he adds "From a physicalistic point of view the theorems of ZF and their FIN(ZF)-counterparts may have the same meaning. Therefore, FIN(ZF) […] eliminates ideal (infinite) objects from the proofs of properties of concrete (finite) objects." ([30]: 59).

[25]) which has some affinities to Mycielski's, adds that FIN(ZF) is a *counterpart* of ZF which belongs to what he calls "finite mathematics" and does not even presuppose the potential infinite.[3]

Surely, the proof theoretician's formulation "T is finitistically reducible" is not the same as Mycielski's "There is a locally finite theory which is isomorphic to T" or as Lavine's "Each theorem of T has a counterpart in finite mathematics". Thus the claims of the latter do not directly contradict the assessment that ZF fails to be finitistically reducible. Yet, finitistic and locally finite theories and those belonging to finite mathematics on the one side, and the relations of proof-theoretical reducibility and of being isomorphic with (or: being a counterpart of) on the other side, turn out to be so closely connected that the investigations of Mycielski and Lavine can be pulled over into the domain of proof theory—suggesting the result that ZF is also finitistically reducible in the proof theoretician's sense. That proof-theoretical conceptions of finitistic reducibility might give the wrong answer to the question, "Is ZF finitistically reducible?", is the topic of this paper.[4]

2. Finitistic Reducibility Defined

In the introduction, I have simply used "ZF is finitistically reducible" as if this phrase were understood well enough as it stands. "T is finitistically reducible" calls for explication, however. For an answer to this problem, it is advisable to turn to investigations belonging to proof theory.[5]

2.1. Proof-Theoretical Reducibility

In [4, 6, 7], we find the following definition, which seems to be widely accepted by proof theorists:

Definition 1. S is finitistically reducible : $\Longleftrightarrow$ there is a finitary theory T such that S is proof-theoretically reducible to T.

But what about "S is a finitistic theory" and "S is proof-theoretically reducible to T"? For the second phrase, precise definitions have been formulated; but there are several of them which are in use in proof theory. Before I discuss four of them (concentrating

[3] See ([25]: 389): "Each theorem of ordinary mathematics has a natural counterpart in finite mathematics. [. . .] Finite mathematics does not even presuppose any form of the potential infinite."

[4] Thus, it does not contain a discussion whether ZF *is* finitistically reducible and, in particular, does not question that it fails to be so.

[5] Actually, one usually has to turn to the writings of S. Feferman. Apart from these, there are not many proof-theoretical texts which deal explicitly with philosophical themes. Most of the definitions discussed in this paper, in particular, have been formulated by Feferman.

on the version from [4, 6, 7]), let me first introduce some terminology and formal background.

Most of the theories considered in this paper will be formulated in the first-order languages L_{PA} and $L_{(QF\text{-}IA)}$ or in L_{PRA}, the quantifier-free fragment of $L_{(QF\text{-}IA)}$. The vocabulary of L_{PA} contains "$\overline{S}$", "$\overline{+}$", "$\overline{\cdot}$" and "$\overline{0}$", whereas the vocabulary of L_{PRA} (and $L_{(QF\text{-}IA)}$) contains, for each primitive recursive function f, a unique function sign $\overline{f}$, i.e., a "primitive recursive function sign". In each of these languages, there is, for each natural number n, exactly one numeral $\overline{n}$ denoting it in $\mathcal{N}$ (= $\langle \mathbb{N}, S, +, \cdot, 0\rangle$, the standard model of arithmetic). PRA is a theory in L_{PRA} whose axioms are, apart from classical logic in L_{PRA}, the recursion equations for all primitive recursive functions, "$\neg \overline{S}x = \overline{0}$", "$\overline{S}x = \overline{S}y \longrightarrow x = y$"; moreover, PRA is assumed to be closed under the induction rule (formulated with formulae of L_{PRA}). (QF-IA) is the theory that results from PRA through the addition of first-order logic.[6]

A theory is a set of sentences which is deductively closed. The theories T investigated here will be axiomatizable; i.e., for each such T, there will be a recursive set Σ of sentences such that $T = \overline{\Sigma}$. For axiomatizable theories S, $T \ldots$, "σ", "τ"... will be used for Σ^0_0-formulas of L_{PA} representing sets of (Gödel numbers of) axioms for T (with the exception of PA and (QF-IA), where I will take "pa(x)" and "(qf-ia)(x)" for these formulas instead).[7] Given a representation τ of (the set of Gödel-numbers of) a set of axioms of T, the common arithmetizations of "proof in T", "provable in T" and "T is consistent" are:

$$\begin{aligned}
&\mathrm{Proof}_\tau(x, y) :\longleftrightarrow \mathrm{Seq}(x) \wedge y = x_{lh(x)\dot{-}1} \wedge \forall v < lh(x)\ (\mathrm{LogAx}(x_v) \vee \tau(x_v) \vee \\
&\qquad\qquad \exists u w < v (x_w = x_u \dot{\rightarrow} x_v)),\\
&\mathrm{Pr}_\tau(y) :\longleftrightarrow \exists x\ \mathrm{Proof}_\tau(x, y),\\
&\mathrm{Con}_\tau :\longleftrightarrow \neg \mathrm{Pr}_\tau(\overline{\ulcorner \bot \urcorner}).
\end{aligned}$$

[8]

Let L be a language with a recursive set of closed terms and $\mathrm{ClT}(u)$ be a Σ^0_0-formula of L_{PA} representing "u is a closed term of L"; then "y is a closed equation in L" is represented by the Σ^0_0-formula $\mathrm{ClEq}(y)$ of L_{PA} defined by

$$\mathrm{ClEq}(y) :\longleftrightarrow \exists u v \leq y\ (\mathrm{ClT}(u) \wedge \mathrm{ClT}(v) \wedge y = \langle \overline{\ulcorner = \urcorner}, u, v\rangle).$$

---

[6] For theories like Q, $I\Delta_0$ + Exp, $I\Delta_0$ + Supexp, ACA_0, ZF and NBG, which are mentioned only at some points in this paper and serve as mere examples, see [43], [9], [23], [41].

[7] Each of the expressions t which belong to L_{PRA}, $L_{(QF\text{-}IA)}$ or L_{PA} has a Gödel-number $\ulcorner t \urcorner$ with Gödel-numeral $\overline{\ulcorner t \urcorner}$. If α is a k-place formula in L_{PA} and $A \subseteq \omega^k$, then "α is a representation of A" is defined as $\forall n_1, \ldots n_k \in \omega\ (\langle n_1, \ldots, n_k\rangle \in A \iff \mathcal{N} \models \alpha(\overline{n}_1, \ldots, \overline{n}_k))$.
For the arithmetical hierarchy, see [9], [20].

[8] The "syntactic" metamathematical expressions "x is a sequence", "the length of (sequence) x", "x is the negation of y", "x is the conditional of y and z", "x is a formula of L_{PRA}" are assumed to be represented by Σ^0_0-formulas of L_{PA}—for which I write "Seq(x)", "$lh(x)$", "$x = \dot{\neg} y$", "$x = y \dot{\rightarrow} z$", "l[pra]($x$)"—in the usual way (see, e.g. [9] and [20]; for the "dot-notation", see also [3]).

Let g be a primitive recursive function with index e; then

$$\sigma\rho_{\mathrm{p}}[e]\tau \quad :\longleftrightarrow \quad \forall x y\,(\mathrm{Proof}_{\sigma}(y,x) \wedge \mathrm{ClEq}(x) \rightarrow \mathrm{Proof}_{\tau}([e](y),x)).$$

Definition 2 (see [4, 6, 7]). $S\ \overline{\rho_{\mathrm{T}}}\ T :\Longleftrightarrow$ S is (non-uniformly) proof-theoretically reducible to $T :\Longleftrightarrow$ there is a primitive recursive function g with index e[9] such that

(a) $\mathcal{N} \models \sigma\rho_{\mathrm{p}}[e]\tau$, and

(b) $T \vdash \sigma\rho_{\mathrm{p}}[e]\tau$.

The problem with proof-theoretical reducibility as it is explained in Definition 2 is its lack of transitivity: see [31] for an example of theories S, T, U such that $S\ \overline{\rho_{\mathrm{T}}}\ T$ and $T\ \overline{\rho_{\mathrm{U}}}\ U$, but not $S\ \overline{\rho_{\mathrm{U}}}\ U$. Of course, this difficulty vanishes if, instead of the varying theories T mentioned in (b), there is just one uniformly chosen theory U *in which* "$\sigma\rho_{\mathrm{p}}[e]\tau$" is proved. Quite often, PRA is adopted for U; I follow this practice[10] and define the *uniform* variant of proof-theoretical reducibility as follows:

Definition 3 (see [7]). $S\ \overline{\rho_{\mathrm{un}}}\ T :\Longleftrightarrow$ S is (uniformly) proof-theoretically reducible to $T :\Longleftrightarrow$ there is a primitive recursive function g with index e, such that

(a) $\mathcal{N} \models \sigma\rho_{\mathrm{p}}[e]\tau$, and

(b) (QF-IA) $\vdash \sigma\rho_{\mathrm{p}}[e]\tau$.

Proof-theoretical reducibility is also defined as *provable relative consistency*. Thus, let's construe the following as definitions of proof-theoretical reducibility:

Definition 4. S is uniformly provably consistent relative to $T :\Longleftrightarrow$ (QF-IA) $\vdash \mathrm{Con}_{\tau} \rightarrow \mathrm{Con}_{\sigma}$.

Definition 5. S is non-uniformly provably consistent relative to $T :\Longleftrightarrow$ $T \vdash \mathrm{Con}_{\tau} \rightarrow \mathrm{Con}_{\sigma}$.

When these definitions are compared, what are their strengths and weaknesses? First, simply proving the relative consistency of S to T is enough for many proof-theoretical investigations. But the relation between S and a finitistic theory T established by it seems to be too loose to make S finitistically reducible in an intuitively convincing sense.[11] Secondly, provable relative consistency has the advantage of being applicable to theories S, T which are formulated in different, maybe even disjoint, languages. If Definitions 2 and 3 are to make sense, the languages of S and T

[9] This is the definition from [7]. Originally (in [4]), g was assumed to be merely partially recursive.

[10] To be precise, I use (QF-IA). For the purpose at hand, the difference between PRA, (QF-IA) and $\mathrm{I}\Sigma_1$ is negligible from a theoretical viewpoint: for (QF-IA) and $\mathrm{I}\Sigma_1$ are conservative extensions of PRA, and (QF-IA) and $\mathrm{I}\Sigma_1$ prove the same Π^0_2-sentences. Practically, the choice of (QF-IA) is preferable because its language contains both quantifiers and all primitive recursive function signs.

[11] This judgement may be too hasty; cf. the Theorem.

mentioned in them should at least share the closed equations. Thirdly, there is the distinction between the uniform and the non-uniform versions: in cases of the latter, it must be possible to formalize (perhaps restricted versions of) "x is a proof for y in A" in the language of the theory T in which proofs are carried out. Thus, the theory of groups, for example, although being a *subtheory* of the theory of abelian groups, is hardly non-uniformly proof-theoretically reducible to the latter.

In sum, the uniform versions seem to be superior. We have, however, the interesting and somewhat surprising situation that for many of the theories usually considered proof-theoretical reducibility as defined by Definitions 2, 3 and 4 amounts to the same. More explicitly:

Theorem.[12] *Let L be an extension of L_{PA}, and let S, T be consistent recursively enumerable theories formulated in L with Σ^0_0-formulas (in L_{PA}) σ, τ representing axiom-sets of S, T, such that $\mathrm{Q} \subseteq S$ and $I\Sigma_1 \subseteq T$. Then*

$$S \overline{\rho_{\mathrm{T}}}\, T,\ S \overline{\rho_{\mathrm{un}}}\, T \text{ and "}S \text{ is uniformly provably consistent relative to } T\text{"}$$

are equivalent with each other.[13]

There still remains the question of why (QF-IA) should be taken for U, the "uniform proving theory". Certainly, this choice is admissible: for it is the purpose of a proof of "$\sigma\rho_{\mathrm{p}}[e]\tau$" in U to guarantee that "$\sigma\rho_{\mathrm{p}}[e]\tau$" actually holds; and in order to fulfill this task, U has to be sound (for Π^0_1-sentences, at least). The consistency and soundness of PA seems to be as much beyond doubt as that of (QF-IA), however. Now, (QF-IA) may be regarded as distinguished in virtue of being finitistic or finitistically reducible; but I think this view is problematic (see below). Finally, why not prefer, for example, $\mathrm{I}\Delta_0 + \mathrm{Exp}$ to $\mathrm{I}\Sigma_1$ when one has finitistic demands on U?

Even if strong conceptual reasons for the choice of (QF-IA) for U are hard to find, there are still some which speak for its adoption that may be called "pragmatic". On the one hand, if PA is taken for U instead of (QF-IA), too many theories would become indiscernible with respect to proof-theoretical reducibility. The reason is that PA proves the outright consistency—whence the relative consistency—of "weak" theories, such as, e.g., all finitely axiomatizable subtheories of PA. But the resulting proof-theoretical reducibility of, say, $\mathrm{I}\Sigma_{1000}$ to propositional logic is certainly not intended. On the other hand, theories which are (intuitively) considerably weaker than (QF-IA)—like Q or $\mathrm{I}\Delta_0 + \mathrm{Exp}$—are too weak to do their job. Thus, consider the pairs ACA_0, PA and NBG, ZF: that ACA_0 (resp. NBG) is proof-theoretically reducible to PA (resp. ZF) is commonly taken to be a paradigmatic case for this intertheoretical relation.[14] If proof-theoretical reducibility were defined as above, but with $\mathrm{I}\Delta_0 + \mathrm{Exp}$

[12] See [7] and [31]. A similar theorem holds if L is assumed to be an extension of $\mathrm{L}_{\mathrm{PRA}}$ and "$\mathrm{I}\Sigma_1$" is replaced by "(QF-IA)".

[13] In particular, $\overline{\rho_{\mathrm{T}}}$ is transitive for these theories. This is not so, however, for non-uniformly provable relative consistency, which fails to be transitive for a wide range of theories, see [31].

[14] Note that in both cases, we do not have relative interpretability results.

in place of (QF-IA), however, these pairs of theories would cease to be examples for proof-theoretical reducibility; for P. Pudlák has shown $\mathrm{I}\Delta_0 + \mathrm{Exp} \nvdash \mathrm{Con}_{\mathrm{zf}} \rightarrow \mathrm{Con}_{\mathrm{nbg}}$ (see [37]; an analoguous result holds for ACA_0 and PA). Theories between $\mathrm{I}\Delta_0 + \mathrm{Exp}$ and $\mathrm{I}\Sigma_1$ could remain as reasonable candidates for U: $\mathrm{I}\Delta_0$ + Supexp, for example, does prove "$\mathrm{Con}_{\mathrm{pa}} \rightarrow \mathrm{Con}_{\mathrm{aca}_0}$"; but since it also proves "$\mathrm{Con}_{\mathrm{q}}$" (see [9]), Q will be proof-theoretically reducible to propositional logic in this case.[15]

At this point, I will not make a decision in favour of or against any of the definitions 2 to 4. Rather, I will deal with each of them as long as it makes sense, making it clear in each particular case which one is under investigation.

2.2. Finitistic Theories: Background

From the foregoing subsection's considerations, it is apparent that "S is finitistically reducible" is precisely defined if "T is a finitistic theory" is. But it is quite difficult to give a convincing explication of the latter. In the remainder of this section, I will deal with several suggestions for carrying out this task. To start with, let me put together some observations regarding the current status of the interpretation of "S is a finitary theory" in proof theory.

(a) "PRA is a finitistic theory"—in short: "FIN-PRA"—is generally accepted.

(b) Often, [44] is cited as providing reason for the assessment that PRA is a finitistic theory.[16]

(c) Sometimes D. Hilbert's texts are taken to express FIN-PRA.[17]

(d) No general and precise definition of "S is a finitary theory" has been agreed upon.

In principle, it is easy to close the gap mentioned in (d) and formulate a definition of "S is a finitary theory" which supports FIN-PRA: one simply takes

(D?) $\qquad S \text{ is a finitistic theory} : \Longleftrightarrow \ S = \mathrm{PRA}.$[18]

On the one hand, such a definition has, to the best of my knowledge, not been suggested in the relevant literature. On the other hand, apart from PRA and PA (or rather the set of its Π^0_2-theorems; see [22]), I do not know of any theory that explicitly has been claimed to be finitary or the limit of finitary reasoning.[19] Be this as it may: I regard (D?) as quite unsatisfactory. Not that definitions by enumeration are something

[15] Thus, uniform proof-theoretical reducibility has its problems, too. In light of these considerations, it should be interesting to find theories which are weak in proving consistency claims yet strong in proving relative consistency claims. One should be aware, however, that this goal is not motivated by basic conceptual considerations.

[16] See, e.g., [2], [19], [36] and, in particular, [40].

[17] For a recent example, see [36].

[18] Or instead, one could, for a finite list of theories $T_1, \ldots, T_k$, choose the adjunction of the equations "$S = T_i$" as the *definiens*.

which is not admissible;[20] what is missing here is rather a *conceptual analysis* that would support (D?) or its variant. After all, we are interested not in a mere stipulation, but in an *explication* of "S is a finitistic theory". And in order to determine whether a suggested *explicans* is adequate for the *explicandum*, some pretheoretic, preformal understanding of the latter has to be developed.

In the case of (D?), there is a further, more specific problem: it violates a principle which—in light of the identification of finitistic theories with those avoiding assumptions of infinity—seems quite plausible:

(P1) If S is a finitary theory and $T \subseteq S$, then T is a finitary theory.[21]

Now, the question emerges as to whose pretheoretic, preformal understanding it should be that counts for determining "S is a finitistic theory". I take it that the first answer is Hilbert, the father of finitism. Yet, I think one should be reluctant in ascribing the view to him that PRA is a finitistic theory. To begin with, Hilbert simply did not give a precise definition of "S is a finitary theory". The vagueness of his writings on conceptual themes notwithstanding, it seems quite clear, however, that both his general conception of metamathematics and his understanding of *finitism* changed over the years. Thus, I agree that in [17] and [18], PRA is accepted as being finitary (see [39] for details); in [18] the same may hold even for PA. But I doubt that this is so in Hilbert's papers from the 1920s:[22] unrestricted variables (whether free or bound) were not allowed in finitistic reasoning then.[23] When it comes to a proper understanding and evaluation of Hilbert's program, I think that particular attention should be paid to this period, however. For it seems that its presentation in [17] and, particulary, in [18] was influenced by P. Bernays' view of the foundations of mathematics and was, in part, an attempt to avoid the consequences that Gödel's Incompleteness Theorems seemed to have for the original form of Hilbert's program.

If, however, it is one's goal to develop a convincing conception of what it is to be a finitary theory, it may be reasoned as follows: since Hilbert's own formulation of his program is vague and precise renderings of it fall prey to Gödel's Incompleteness Theorems and similar results,[24] there is no way around making corrections to Hilbert's original claims—and develop one's own, if possible superior, variant of finitism. Yet,

[19] In [44], W. W. Tait not only endorses FIN-PRA, but also the claim that the universal closure of each theorem of PRA is finitistically provable. I do not think, though, that one should therefore ascribe the acceptance of (D?) to him. Tait's claim that finitistic means do not go beyond what can be proved in PRA does not preclude subtheories of PRA from being finitistic and does not fix the extension of "finitistic theory" when dealing with theories which are not formulated in arithmetical languages.—When [44] gets interpreted in this way, it does not contain a *general* definition of "S is a finitistic theory", however.

[20] Although they hardly provide the *general conditions* I am asking for.

[21] Perhaps T should be restricted to recursively enumerable theories in (P1).

[22] For example, the passages from Hilbert's texts cited by B. Rolf [38] and G. Kreisel [21] in support of their analyses of his work are mainly taken from [17] and [18]. But these authors do not distinguish between Hilbert's program of the 1920s and that of [17, 18] and, accordingly, view their interpretations as being correct for the former period, too.

[23] For details of this interpretation of Hilbert's program, see [33].

[24] See, however, [34].

what is the limit to this sort of deviation from Hilbert's original? What are the criteria for calling a position, a procedure or a theory "finitistic"? It may be "superior" in virtue of being precisely defined, free of logical and metamathematical flaws – but its philosophical motivation may be weaker or lost. With these remarks, the question of the motivation of Hilbert's own program is also addressed. As I see it, Hilbert style metamathematics M has, briefly put, the following role. In M, the consistency of theories T of formalized mathematics has to be established.[25] This is done by showing in M that no proof in T can end with, say, "$\overline{0} \neq \overline{0}$". In order to fulfill this task, M must be absolutely reliable—it must contain only "absolute truths" (see [10]: 35 in [16]). Such reliability is supposed to be guaranteed by the fact that, given an arbitrary finite string s of concrete objects (e.g., signs *qua* tokens), M merely contains reports (like *protocol sentences*) about s and the results of simple manipulations on s—such as extending, shortening and concatenation. This is where *finitistic restrictions* are imposed on M; in short, their role is to secure the reliability of M.[26]

It is not the aim of this paper to formulate an adequate and still precise elaboration of Hilbert's program as sketched. In particular, I will not question the truth of FIN-PRA in it; rather, I am interested in the informal conceptions of finitism underlying this claim. Given (b), a discussion of [44] should be the first natural step at this point; in fact, I have done this elsewhere together with M. Schirn.[27] Here, I will concentrate on analyses of "S is a finitary theory" which are conceptually different from Tait's—in particular those suggested by Feferman's, Mycielski's and Lavine's work—and see whether they can be used to formulate general definitions of "S is a finitary theory".[28]

---

[25] Hilbert's program is sometimes taken to include a *conservativity program* as a central component, see, e.g., [21], [8], [42]. That is, given a finitistic theory S and an arbitrary formal mathematical theory T, it is said that Hilbert wanted to show:

If ψ is a real sentence and $T \vdash \psi$, then $S \vdash \psi$.

If this interpretation is supported by quotations from Hilbert's own texts, merely one passage is usually cited (from [11], see [46]: 376): "One can claim, however, that [the science of mathematics] is an apparatus that must always yield correct numerical equations when applied to integers". In fact, a similar claim is made in [12] (see [46]: 471) and perhaps another one in [13]. But these are less than 10 lines on the theme of conservativity in the more than 50 pages of [11, 12, 13]. And in the earlier papers collected in [16], there is not a single sentence where Hilbert addresses this topic. In all of these texts, the aim of showing the consistency of T by finitistic means is always in the foreground of Hilbert's interests.

Thus, Hilbert's papers from the 1920s do not contain enough textual evidence for the attribution of a conservativity program to Hilbert. Let me add that the equivalence of T's consistency and T's soundness for Π^0_1-sentences (for Σ^0_1-complete theories T) does not constitute an argument for the assessment that Hilbert was, after all, pursuing a conservativity program since he was interested in a consistency program. For the predicate "being interested in showing ψ" is, like other contexts dealing with propositional attitudes, *hyperintensional*: even from "a is interested in (has the goal of, program of) showing ψ" and "ψ is *logically* equivalent to φ", "a is interested in (has the goal of, program of) showing φ" does not follow.

[26] In some papers on finitism, justice is done only to the goal of reliability, whereas the demand for finiteness is simply ignored. An example is [22], where it is suggested that each sentence obtainable by the iterated addition of certain reflection rules to PRA should be taken to be finitistically provable.

[27] See [35]. Actually, we do not share the common acceptance of [44]. In short, our complaint is that if Tait had succeeded in showing that the principles put forward by him in [44] were finitistically correct, he would obtain much more than PRA as the limit of the finitistically admissible.

2.3. Finitistic Theories: Feferman

In [6]: 8, Feferman writes: "[...] the use of classical quantificational logic in *any* system containing the base axioms (1) $x' \neq 0$ and (2) $x' = y' \rightarrow x = y$ [...] implicitly requires assumption of the completed countable infinite." In a more elaborated version of this assertion appearing in [4]: 374, it is said that any such system or theory "[...] implicitly requires for its justification an appeal at least to the *completed infinite totality of natural numbers*, in other words to the *countable infinite*." Moreover, we find in ([4]: 374): "On the face of it, we have the following [...] PRA is justified on finitary grounds" and in ([6]: 8): "It is generally acknowledged that PRA is a finitarily justified system, or to be more precise, that each theorem of PRA is finitarily justified."

Yet, why should it be the case that each theorem of PRA is finitarily justified? Feferman does not explicitly give an answer, but in light of the passage just quoted a first guess may be that this is so because for each theorem ψ of PRA, it is *not* necessary to appeal to the "completed infinite totality of natural numbers". This in turn could be made precise by the claim

(AF-PRA) Each formula ψ (of L_{PRA}) provable in PRA has a finite model.

But (AF-PRA) is false. Already the conjunction of the formulas mentioned in the above citation from [6],

$$x' \neq 0 \ \wedge \ (x' = y' \rightarrow x = y) \tag{1}$$

cannot be satisfied in a finite model. Note that in this respect there is no difference between (1) and its universal closure (formulated in $L_{(QF\text{-}IA)}$). Now, an alternative to (AF-PRA) worth considering is its version for closed formulas only:

(CAF-PRA) Each sentence ψ (of L_{PRA}) provable in PRA has a finite model.

Contrary to (AF-PRA), (CAF-PRA) is indeed true. Moreover, we have a relevant difference between PRA and (QF-IA) (or $I\Sigma_1$) at this point: the universal closure of (1) provides an example for a closed theorem of (QF-IA) without a finite model.

Generalizing from the case of PRA to arbitrary theories in a (first-order) language L, one obtains for a sentence ψ of L

(D1) ψ is finitarily justified : $\Longleftrightarrow$ ψ has a finite model.

Now, Feferman called PRA a "finitarily justified system" because each of its theorems is finitarily justified, cf. also [44]. Since that assertion is correct under interpretation (CAF-PRA), but not under (AF-PRA), I suggest to abstract the following definition

[28] In these approaches, the main idea is *finite models for finite systems*. The other two types of precise explications of "finitistic theory" I am aware of employ versions of the ω-rule (see most notably [44]) and progressions of theories (see [22]). In comparison, the former type has the advantage of resting on claims of Hilbert: see [14, 15]; similar suggestions made by him as regards the use of progressions are not known to me.

from the case of PRA:

(D2) T is a finitistic theory $:\Longleftrightarrow$ Each sentence ψ (of L_T) provable in T is finitarily justified.

Note that by (D1), the *definiens* of (D2) is just

(CAF-T) Each sentence ψ (of L_T) provable in T has a finite model

Furthermore, (D2) clearly has (P1) as a consequence.

Let me mention that Feferman has not explicitly stated that all theories T which satisfy (CAF-T) are finitistic. But since I do not find any other general explanation of "T is a finitistic theory" in his (or, with the exception of Tait, in other proof theoreticians') writings, I will nonetheless go on with a discussion of the above definitions. Now, with (D1) I am in agreement; for it is quite plausible to explain "ψ does not make assumptions of infinity" by "ψ has finite models". Choosing "ψ has only finite models" as an *explicans* instead should be rejected, however. Such an explanation would lead to the inacceptable result that, for example, the logical truths of first order-logic would make assumptions of infinity (since they have infinite models). Moreover, if (P1) is accepted, this would imply that no finitely axiomatized theory formulated in a first-order language could be finitistic. Thus, a finitarily justified sentence will also have infinite models.

(D2), however, seems to be quite problematic. The criticism which immediately comes to mind is that there are theories declared to be finitistic by (D2) which have no finite models at all. PRA is an example, and others will play a major role below (see section 3). Furthermore, there are theories of this sort which are even, in some sense, *paradigmatic* for making assumptions of infinity: take, for example,

$T_1 := \overline{\{\exists_{\geq n} \mid n \in \omega\}}$—formulated in the first-order language L_1 containing only the identity sign, where $\exists_{\geq n}$ shall be a sentence satisfied only in models with at least n elements; and

$T_2 :=$ the first-order theory of atomic Boolean algebras extended by $\{\exists_{\geq n} At \mid n \in \omega\}$—formulated in a first-order language L_2 with "$\leq$" and "$=$", where $\exists_{\geq n} At$ shall be a sentence satisfied only in models with at least n atoms.

T_1 and T_2 are both finitistic in the sense of (D2), but neither of them has finite models. In addition, T_1 is *the* theory in L_1 that has only infinite models: if there is any theory in L_1 having only infinite models, it has to contain T_1, whence it must be identical with T_1 (since T_1 is complete in its language). Similarly, T_2 is *the* theory in L_2 having only infinite models which extends the first-order theory of atomic Boolean algebras.

Furthermore, in my opinion it is quite implausible that PRA should be "finitarily justified", whereas (QF-IA) should require the "completed infinite totality of natural numbers". Certainly, the theorems of first-order predicate logic are true in each (nonempty) domain: they are not *transfinite* axioms (to use Hilbert's unhappy terminology) in that they would require infinite domains for their truth. Since (QF-IA) results from PRA simply by superimposing first-order logic on it and, by assumption,

PRA also does not require the "completed infinite totality of natural numbers", it seems odd that (QF-IA) should do this. It may also seem hardly tenable that PRA should be finitistic or potentially infinitistic, but theories such as the theory of dense linear orderings without endpoints (DLO) and Q—which intuitively and from the point of view both of relative interpretability and of proof-theoretical reducibility are much weaker than PRA—fail to be so (since (CAF-DLO) and (CAF-Q) are false). One should be aware of the fact, however, that all of the mentioned theories are finitistically reducible.

Of course, in these considerations I merely report what I view as tensions between a first, crude intuitive understanding of "S is finitistic" and its definition (D2). Whether or not these are relevant to Feferman's ideas is hard to decide, for no analysis of, for example, "implicitly requires assumption of the completed countable" is forthcoming in [4, 6]. To some extent, the criticism put forward here is terminological; accordingly, it may be countered by a change of terminology. Since in the passages quoted from [4, 6] (and some others) there are indications to the theme of the *potential infinite*, Feferman might accept a definition such as

(D2′) T is a potentially infinitistic theory : $\Longleftrightarrow$ Each sentence ψ (of L_T) provable in T is finitarily justified.[29]

I will, however, postpone a more thorough discussion of how to understand "potential infinity" and a further criticism of (CAF-T) to sections 2.5 and 3.

2.4. Finitistic Theories: Mycielski and Lavine

In [29] and [30], Mycielski has presented a method of obtaining for each recursively enumerable first-order theory T a *locally finite* theory FIN(T), as he calls it, such that T is translatable (in a specific sense) into FIN(T). "T is locally finite" is explained through "every finite part of T has a finite model" ([30]: 59). Thus, T is locally finite if and only if (AF-T) or (CAF-T) holds.[30]

With respect to FIN(T) and the translations $\mathfrak{I}$ from T to FIN(T) envisaged by Mycielski, suffice it to point out:

– the relevant axioms of FIN(T) result from the axioms of T by relativizing their quantifiers to predicates Ω_i ($i \in N$) which do not belong to L_T—through a procedure called "regular relativization": if the quantifier $Q'x$ (i.e., $Q' \equiv \forall$ or $Q' \equiv \exists$) occurs in the scope of the quantifier Qy in some formula ψ of L_T, and if Qy gets relativized to Ω_i and $Q'x$ to Ω_j, then $i < j$ has to hold;

[29] In this case, PRA would be declared to be a potentially infinitistic theory rather than a finitistic one, though.

[30] Actually, although Mycielski's definition is ambiguous with regard to these two readings, I think that the examples he deals with suggest that he prefers (AF-T); in this case, PRA would turn out to be not locally finite.

– each $\mathcal{I}$ is similar to a relative interpretation: it is a function mapping a formula ψ from L_T to a regular relativization of it; thus, $\mathcal{I}$ is the identity function on atomic formulas, it commutes with the sentential operators and relativizes quantifiers as mentioned;

– each such $\mathcal{I}$ maps T into $\mathrm{FIN}(T)$.

As regards the relation between locally finite theories and what Mycielski takes to be finitistic ones, there are two remarks in [29, 30] containing the predicate "finitistic". The first (see section 4.5 from [29]), being a comment on the locally finite theory FIN, expresses a position that seems to be close to the one described in the previous section: it is compatible, for instance, with (D2) and (D2′). Mycielski's presentation is, however, too sketchy to determine what exactly he takes to be the relation between finitistic theories, locally finite ones and those resting on the potential infinite.[31]

The second passage is from [30]: 62: "Both PA and PA + (*) [i.e., PA + RFN[pa]] deal only with finite objects and stem only from our experience with finite objects. Therefore (P_1) and (P_2) constitute a finitistic reduction of the problem of local finiteness of $\mathrm{FIN}(A)$ to the problem of consistency of A". The formulas (P_1) and (P_2) mentioned in this citation are "$\mathrm{PA} \vdash \forall F\ [\mathrm{PA} \vdash (\mathrm{Con}(A) \rightarrow (F \text{ has a finite model}))]$" and "$\mathrm{PA} + (*) \vdash (\mathrm{Con}(A) \rightarrow (F \text{ is locally finite}))$". If what Mycielski writes is understood as implying

PA and PA + RFN[pa] are finitistic theories,

it is unacceptable for the proof theorist.[32] Moreover, PA and PA + RFN[pa] are far from being locally finite and it is hard to see why they should rest merely on the potential infinite.[33] Indeed, in saying that they deal only with finite objects and stem from our experience with finite objects, Mycielski puts forward a new type of argument for the supposed finitistic character of these theories. That PA and PA + RFN[pa] do so is a quite plausible assessment; but it is not clear that this provides a reason for calling them "finitistic". After all, there are models of ZF containing only "the" natural numbers, i.e., finite objects; furthermore, not all models of PA or PA + RFN[pa] contain only "finite" numbers.[34] Surely, one need not cling to model-theory. Yet, even intuitively, it seems that $\mathrm{PA} + \mathrm{Con}_{\mathrm{zf}}$ *deals only with finite objects* (if ZF is consistent); and claiming that it *does not stem from our experience with finite objects* is so vague that it hardly provides a reason to deny that $\mathrm{PA} + \mathrm{Con}_{\mathrm{zf}}$ is finitistic.[35]

[31] In addition, such expressions as "… is syntactically isomorphic to …" and, in particular, "… has the same meaning as … from a physicalistic point of view" also remain unexplained and undiscussed in [29, 30]. This makes it difficult to understand and evaluate some of Mycielski's (supposedly) central claims—such as *FIN[ZF] is syntactically isomorphic to ZF* and *From a physicalistic point of view the theorems of ZF and their FIN[ZF]-counterparts may have the same meaning*.

[32] Lavine claims that Mycielski accepts PA as being finitistic; see [25]: 415.

[33] In [30], the expression "potentially infinite" is only applied to theories.

[34] Talk about *the standard model* should be problematic for such a devoted Anti-Platonist as Mycielski.

[35] If $\mathrm{PA} + \mathrm{Con}_{\mathrm{zf}}$ were finitistic, ZF would not only be finitistically reducible in the sense of Definition 1, but even be relatively interpretable in a finitistic theory; for a related reasoning, see section 3.

Let's turn then to Lavine, who in [24] and [25], by applying some of the metamathematical results of Mycielski, has developed an approach to the foundations of mathematics which has some similarities with Mycielski's, but is much more elaborated in philosophical matters. In discussing Lavine's contribution, let me focus on its distinguishing features.

To start with, Lavine claims that both FIN(ZF) and PRA belong to what he calls *finite mathematics*, see [24]. He explains "finite mathematics" as follows: "Only finite entities are employed, and all quantifiers within finite mathematics have finite ranges" ([25]: 389); "our commitments within a theory of finite mathematics are consistent with a statement of the form 'there exist at most n objects' for some natural number n" ([25]: 394); and "The point of finite mathematics is not that its theories have no infinite models, but rather that its theories do have finite models" ([25]: 398). These assertions fit (D1); but they make it hard to understand how FIN(ZF) and PRA could be theories of finite mathematics: for these theories do not have finite models. At this point, however, a specific conception of what a theory is enters the picture. Usually, one regards a theory T as an infinite set of formulas—the theorems derivable from a set of axioms of T, say. And, for example, "FIN(ZF)" and "PRA" are taken to be terms denoting such infinite sets. Not so Lavine: he views theories as something developing—perhaps growing.[36] Furthermore, he should be interpreted as construing, e.g., "PRA" not so much as a term—which denotes or purports to denote a set of formulas—but rather as a one-place formula "PRA(x)" (or: as a predicate "PRA") which is true of these formulas, cf. [25]: 416. Since whenever "PRA(x)" is applied to formulas, it is applied to actually presented ones, it is plausible to assume that at each of its applications, "PRA(x)" is true of only finitely many formulas—of finitely many PRA-theorems (if these words are understood as usual).[37] Eventually, the claim "FIN(ZF) and PRA belong to finite mathematics" is thus explained as "Each finite subtheory of FIN(ZF) or PRA belongs to finite mathematics". But this, in turn, amounts to nothing more than "Each finite subtheory of FIN(ZF) or PRA has a finite model", if at this stage, "belongs to finite mathematics" is explained as quoted above. Thus, we are back where we started: "T belongs to finite mathematics" gets defined as (CAF-T) or perhaps as (AF-T).

At this point, the failure of (AF-PRA) may be considered a difficulty for the statement that PRA belongs to finite mathematics. Lavine could be perhaps be content with (CAF-PRA); but as I understand him, he has a bolder answer to this challenge: he views PRA as a *schematic* theory. Now, one problem is that, given a common understanding of "schematic theory", PRA is not one. Typical examples for schematic theories are PA and ZF: their (usual) axiom-systems encompass *axiom schemata*—the induction schema and the axiom schemata of separation and replacement or collection.

[36] Of course, this is an old idea, to be found e.g. in the philosophy of Leśniewski; moreover, it was also suggested by Mycielski ([30]: 59): "The concept of a locally finite theory is interesting only if the theory itself is regarded as a process of formation of axioms and theorems and not as an actually infinite object. But this internal view is of course the most natural view of any theory in which one proves theorems."

[37] Similarly for FIN(ZF).

PRA, however, simply has infinitely many axioms, all of which are formulated with free variables, and no axiom schema.[38]

Maybe Lavine takes the theory PRA to be the deductive closure of all the *closed instances* of the usual PRA axioms.[39] In this case, PRA would be a subset of the deductive closure of PRA^C which is defined through

$$\mathrm{PRA}^C := \{\psi \mid \psi \text{ sentence in } \mathrm{L_{PRA}} \wedge \mathrm{PRA} \vdash \psi\}.$$

Certainly, the theory PRA—as this term is usually understood—is not the same set as PRA^C (in distinction to PRA, PRA^C has no open formulas as elements); but equally trivially, PRA and PRA^C contain and prove the same sentences. For the question whether PRA and PRA^C are the same theory, however, it is relevant whether PRA and PRA^C prove the same *formulas* of $\mathrm{L_{PRA}}$. Now if this were the case, because of $\mathrm{PRA} \vdash (1)$, $\mathrm{PRA}^C \vdash (1)$ would hold. Since (1) is just one formula, there would be a finite subset E of PRA^C such that $E \vdash (1)$. But then, by (CAF-PRA), E has a finite model, whence (1) has a finite model. But this cannot be the case. If Lavine interprets "PRA" as PRA^C or its deductive closure, he is simply not dealing with the theory PRA as it is usually construed.

When understood in Lavine's sense, theories may be regarded as examples for "the potential infinite". Now, the topic of the *potential infinite* is mentioned at several places in my paper, but a discussion of it is still missing in it. Of course, I do not have the place to close this gap; but in the following subsection, I will address some of the difficulties that explications of "x is potentially infinite" and "T assumes the potential infinite" have to face.

2.5. The "Potential Infinite"

The statements "a is finite" and "a is infinite" have found convincing and precise set-theoretic explications.[40] This is not the case for phrases of the sort "x is potentially infinite" and "T assumes the potential infinite" (or "T rests on the potentially infinite"). It is granted that there are stereotypical assertions which could serve as explanations for them, many of them taking Aristotle's "In general, the infinite is in virtue of one thing's constantly being taken after another—each thing taken is finite, but it is always one followed by another" (see [1]: III 6.206a) as their starting point. But formulations such as these are far from being precise definitions. That this is so is all the worse since the distinction between the potential and the actual infinite has often been regarded as being of fundamental importance and has regularly been employed as the motivational basis for further philosophical work. "x is potentially infinite" and

[38] Actually, this also holds when the explications of "T is a schematic theory" presented in [5] are applied. Let me add that I have found no definition of this phrase in Lavine's texts.

[39] That's not exactly what he does, but (I think) it amounts to the same; cf. [24]: VI.3/4 and IX.6.

[40] There are several of them which are S-provably equivalent to each other if (the set theory) S is not too weak.

"*T* assumes the potential infinite" are much too unclear, however, to be considered acceptable end-points of an explication; rather, they themselves are much in need of one.

In what follows, I will simply take for granted recent writings of philosophers and logicians who have some sympathy for the topic of the potential infinite; and I will use them as the material for comments and further suppositions.[41] Because of his quite explicit assertions, I begin with P. Lorenzen. In [26], he presents a pair of rules as a means to obtain infinitely many objects

"(1) Start with I.

(2) If one has reached x, one should append I to x."

and continues with "Now we can say at once that applying these rules, infinitely many numbers are possible: for each number x, x I has still to be constructed. One must be aware that here the mere *possibility* is claimed [. . .] The rule *makes it possible* to construct xI, if x has already been constructed [. . .] Yet, to claim that infinitely many such numbers be real, i.e., had in reality been constructed following these rules—this would, of course, be wrong".[42]

First, it should be noted that this quotation contains no definition—be it precise or not—of "*x* is potentially infinite" and "*T* assumes the potential infinite" at all. Rather, *theses* are formulated in it which, generalizing from Lorenzen's own case, the *potential-infinitist*—as I will call one who accepts assumption of the potential infinite, but rejects the assumption of the actual infinite—should accept. Thus, I take it that he proposes

(i) By applying rules (1) and (2), infinitely many natural numbers can be constructed,

(i′) By applying rules (1) and (2), infinitely many natural numbers are possible,

but rejects the claims

(ii) There are infinitely many natural numbers,

(iii) Infinitely many natural numbers have been constructed.

[41] I have to admit that, given the modern conception of sets in classical logic, I have never understood "potentially infinite" in a way which would give the potential infinite any use—a place of its own different from the finite and the infinite. It seems to me that merely in intuitionistic frameworks the predicate "potentially infinite" may find a reasonable rational reconstruction. Thus, in dealing with this discourse I feel as if I were a (Quinean) linguist in the process of developing a radical translation from a foreign language into his mother tongue.

[42] My translation; the original is "(1) Man fange mit I an. (2) Ist man zu x gelangt, so füge man noch x I an" and "Jetzt können wir sofort sagen, daß nach diesen Regeln unendlich viele Zahlen möglich sind: zu jeder Zahl x ist ja noch x I zu konstruieren. Man muß darauf achten, daß hier nur die *Möglichkeit* behauptet wird [. . .] Die Regel *ermöglicht* xI zu konstruieren, wenn x schon konstruiert ist. [. . .] Dagegen zu behaupten, daß unendlich viele solche Zahlen wirklich seien, also wirklich nach diesen Regeln konstruiert seien – das wäre natürlich falsch" ([26]: 4).

Actually, it is quite common that theses containing the expression "potentially infinite" are stated instead of definitions for "x is potentially infinite". Thus, because of the difficulties in finding such definitions, an approach such as Lorenzen's has its advantages. In fact, it is neither easy to extract the sought-after definitions from the theses nor do the theses depend on prior definitions.

Second, (i) and (iii) are apparently assertions about what is possible for—what can be done by—human beings; and the expression "is possible" occurring in (i′) should probably also be understood in this sense, making (i) and (i′) equivalent with each other. Under this reading, the rejection of (iii) is quite plausible; and whether the negation of (i) is true depends on the exact interpretation of the word "can" occurring in it.[43] If "x is possible" is not interpreted as such a constructibility claim, however, (i) and (i′) need not be equivalent; for it is certainly not evident that those entities which can be constructed by human beings and those which are possible are one and the same objects. Given the reigning conception of the ontology of mathematical objects, in particular, such an identification is bound to fail there—if it makes sense at all: numbers, for example, are possible; but not a single one has ever been or can be constructed.[44] Moreover, in this case the denial of (iii) does not give a reason for the denial of (ii).

For an example for what may be taken to be a definition of "x is potentially infinite", let's turn to an article of J. Thomson. Following Aristotle, he considers some divisible object and distinguishes between "whatever finite number n you take, it is possible for the thing to have been divided into more than n parts" and "it is possible that the thing could have been divided into more than any finite number of parts" ([45]: 186). These statements suggest the following pair of definitions (let's write "$\Diamond\psi$" for "it is possible that ψ"):[45]

(D3) a is potentially infinite : $\Longleftrightarrow\ \forall n\ (n \in \mathbb{N} \rightarrow \Diamond(a \text{ has } n \text{ parts}))$[46]

(D4) a is actually infinite : $\Longleftrightarrow\ \exists z(\Diamond(a \text{ has } z \text{ parts}) \wedge \forall n(n \in \mathbb{N} \rightarrow n \leq z))$.

I think that, historically, much criticism of the acceptance of actually infinite objects was, in effect, directed not against their very existence, but against treating them as natural or real numbers. Surely, it is unacceptable also from a modern perspective that a natural or real number z can satisfy "$\forall n(n \in \mathbb{N} \rightarrow n \leq z)$)". But only if all the "numbers" taken into consideration are real numbers, the nonexistence of such a z results. That z has to be a natural number is, however, not claimed in (D4).[47] Thus,

[43] Of course, there is the well known general difficulty of interpreting modal terminology.

[44] The proper objects that can be constructed are numerals (*qua* tokens).

[45] Actually, other distributions of modal operators in the *definientia* may be plausible, too. Furthermore, since an object can probably be potentially/actually infinite for reasons other than having many *parts*, implications from right to left seem to be more appropriate than equivalences.

[46] This is similar to [26], but there are two important differences: first, that there are infinitely many natural numbers is presupposed in (D3); second, it is quite implausible to define "a is potentially infinite" by "infinitely many objects could be larger than a": i.e., the acceptance of (D3) as a *definiens* builds on a specific understanding of "part of". Note in this connection that the unmodalized version of (D3) is closely related to the axiom of infinity known from mereology.

the development of the theory of infinite ordinal numbers has saved the adherent of the assumption of the actual infinite—in short: the *actual-infinitist*—from being committed to an inconsistency: by elaborating the *definiens* of (D4) as

$$\exists z\ (z \text{ is an ordinal number } \wedge \Diamond(a \text{ has } z \text{ parts}) \wedge \forall n(n \in \mathbb{N} \rightarrow n \leq z)),$$

he can maintain that there are actually infinite objects.

Can there exist a finite object a such that (D3) holds for it? If a is a concrete object, perhaps—since in this realm, there is change. When it comes to pure mathematical objects a for which "having parts" is meaningful, the *definientia* of (D3), (D4) and "a is infinite" seem to be equivalent, however. This is so, in particular, when "a has z parts" is interpreted as "a has z elements" or as "a has z subsets". For given the common understanding of modal talk as applied to mathematical objects, whatever is the case in this domain is necessarily the case. Thus, if the $\Diamond$'s are erased in (D3) and (D4), the resulting *definiens* of (D3) implies that of (D4): in both cases, a is infinite; and the object z whose existence is asserted in (the modification of) (D4) may be taken to be ω.

This reasoning is connected with a general problem for attempts to explicate "x is potentially infinite". Assume that "x is potentially infinite" is a meaningful phrase, and assume that it is neither equivalent to "x is finite" nor to "x is infinite" (which would make "x is potentially infinite" superfluous). Now, consider an arbitrary a which is potentially infinite. Is a finite or infinite? By classical logic—which is assumed here—one of these alternatives must be the case.[48] Now, if a is finite, one has a result which is even stronger then a being potentially infinite; and if a is infinite, the additional claim that it is also potentially infinite does not save it from being infinite. Thus, in both cases, the claim that a is potentially infinite is irrelevant for someone who cares about distinguishing between the finite and the infinite, e.g., to someone who wants to avoid assumptions of infinity.

So far, I have expressed doubts as to the possibility of a non-trivial explication of "a is potentially infinite". Let me now deal with the theme of the *assumption of the potential infinite*. As I have described him, the potential-infinitist rejects not only the assumption of infinite objects—such as the set of natural numbers—but also the assumption of infinitely many objects. But he tends to accept assertions such as

(iv) Only the finite initial segments of the sequence of natural numbers exist.[49]

and

(v) As the systematic paradigm of the potential infinite, rule-governed non-terminating procedures like "always-counting-on" may be regarded.[50]

[47] It has to be admitted, though, that it is not implausible that a z such that $\Diamond(a \text{ has } z \text{ parts})$ should be a *natural number*.

[48] The equivalence of "x is infinite" with "$\neg$(x is finite)" may be secured by definition alone or by some weak uncontroversial set theoretical principles.

[49] —but not that sequence itself.

On the face of it, however, there is a reference to the *sequence of natural numbers*[51] in (iv) and to the *procedure of always-counting-on* in (v). For, e.g., the predicate "is a finite initial segment of the sequence of natural numbers" must be true of some object a when (iv) is true. In this case, "$\exists z$ (y is a finite initial seqment of $z \wedge z =$ the sequence of natural numbers)" has to be true of a; and this can only be so when the term "the sequence of natural numbers" denotes something—both intuitively and, for instance, from a Quinean point of view. But since the sequence of natural numbers and the procedure of *always-counting-on* are infinite objects, we have assumptions of infinity—of the *actual infinite*[52]—in these examples. It seems that in the very pronouncement of his position, the *potential-infinitist* makes the infinity-assumptions he wanted to avoid.[53]

As I see it, there could be two ways out of this quandary.

(First way: terms vs. predicates) Of course, the *potential-infinitist* should reject the claim that in (iv) we refer to the sequence of all natural numbers. Thus, he could suggest a sentence such as

$$\text{(vi)}\quad \forall n(\text{Natural-number}(n) \rightarrow \exists s(\text{Sequence}(s) \wedge \text{length}(s) = n \wedge \forall i < n\, \text{Natural-number}(s(i))))$$

as a paraphrase of (iv). Note that in (vi) the *term* "the sequence of natural numbers" occurring in the formal paraphrase of (iv) considered earlier has been replaced by the *predicate* "is a natural number". Moreover, the terms remaining in (vi) merely take finite objects as their semantical values. This is also the case for a sentence like (vii) "5 is a natural number"—which intuitively should be acceptable to the potential-infinitist.

In some sense, however, even (vii) makes an assumption of infinity; and so does (vi). For under the assumption of Tarskian (i.e., model theoretic) semantics, if (vii) is true, the predicate "is a natural number" is assigned an extension which is the set of natural numbers (or: a set playing the role of this set in a certain model of a suitable theory), and this simply *is* infinite (given common assumptions). Now, this type of semantics can surely be relinquished in favour of a semantics containing conditions of the sort

"5 is a natural number" is true $\iff$ "is a natural number" is true of 5

instead.[54] Yet, even here, the *predicate* "is a natural number" is still *true of* infinitely many numbers.

[50] This is translated from the article *unendlich*/*Unendlichkeit*; in [28]: 389. Originally, it is "Als systematisches Paradigma der *potentiellen* Unendlichkeit können geregelte, nicht abbrechende Verfahren wie das 'Immer-weiter-Zählen' angesehen werden".

[51] A sequence s is analyzed as a set of ordered pairs $\langle x, y\rangle$ (both of which are natural numbers in our example) such that with $\langle x, y\rangle \in s$ and $u < x$, some v with $\langle u, v\rangle \in s$.

[52] I treat "a is actually infinite" as a synonym of "a is infinite".

[53] See [45]: 186, for a similar view: "The infinite sequence [1, 1/2, 1/4, ...] is a mathematical representation of the continued (potentially infinite) operations of cutting .... But the sequence of winters to come is nonetheless infinite, and in just Cantor's sense".

[54] For details, see [27].

(Second way: linguistic vs. non-linguistic objects) In analogy to (iv), (v) and Lorenzen's claims seem to be committed to procedures and to rules; and one might try to get rid of these commitments in the way just discussed (by employing predicates "is a so-and-so procedure/rule" instead of terms). If, however, the non-terminating procedure of 'always-counting-on' and Lorenzen's rules (1) and (2) are entities, the question turns up of what sort of entities they are. Are they infinite objects in this case? This seems plausible,[55] in particular when rules and procedures are conceived of as functions—as it is typical for the mathematical discourse: in our cases, one would have a function mapping each natural number to its successor—certainly an (actually) infinite object. Of course, it could be responded that procedures and rules can be stated in finitely many words; but this reply is just sloppily formulated. It might mean that those procedures and rules *are* these finite expressions—a position held by F. Waismann with respect to *laws*: "Actually applied are only such sequences for which one knows a specific law of construction. Under such a law a formula has to be understood or a description in words."[56] This construal may fit assertions such as (i) and (i′), for example. Or the above suggestion could mean that those finitely many words express such a procedure or rule (which is closer to (v)). Yet, for the present topic, it does not really matter whether rules are (finite) expressions which apply or whether they are expressed by (finite) expressions; in neither case is there any guarantee that assumptions of infinity are avoided: a finite expression may apply to infinitely many objects, and a procedure or rule expressed by a finite expression need not be a finite objects.

In sum, those two ways out do not really work: in both cases, there are assumptions of infinity—be it of infinite objects or be it of infinitely many objects;[57] be it by interpreting terms or be it by interpreting predicates. Yet, to close this part, let me discuss a final reply to this challenge. This was already addressed in the description of Lavine's conception of theories, which in a nutshell is that "[. . .] schemas do not commit us to the truth of a potential infinity of instances but rather to the truth of the finitely many instances that we actually come to invoke [. . .]" ([24]: 260). Similarly, if a procedure or a rule—such a adding 1—is taken to be applied only finitely many times, only finitely many objects are obtained *de facto*. Furthermore, a predicate like "is a natural number" should not be interpreted as being true of infinitely many numbers, but only as being true of the (finitely many) natural numbers it is applied to.

I do not know what this sketch exactly amounts to and if it can be worked out properly. A formal (Tarskian-style) semantics along its lines is certainly missing, and it is not clear if it can be developed without making assumptions of infinity; but it

[55] Actually, I have assumed this above.

[56] My translation; the original is "Tatsächlich angewendet werden jedenfalls nur Folgen, für die man ein bestimmtes Bildungsgesetz kennt. Unter einem solchen Gesetz ist entweder eine Formel zu verstehen [. . .] oder eine Beschreibung in Worten" ([48]: 119).

[57] I take for granted that in assuming the existence of infinitely many objects one makes an assumption of infinity—even if each of them is finite.

may be declared to be superfluous by the adherents of that approach.[58] One problem I have with it is the following: how can the *potential-infinitist* express that one and the same predicate—like "is a natural number"—is true of or can be true of the finitely many objects it has actually been applied to *and* of the infinitely many objects it could be applied to? Or is it that claims of the sort " 'is a natural number' could be true of infinitely many objects" are withdrawn? In this case, however, I do not see that *any* assumptions of infinity—be they of the potential or the actual infinite—have remained.[59]

To close this subsection, let me quickly deal with explications of phrases of the type "theory T assumes (merely) the potential infinite". On first thought, one will probably tend to formulas containing "x is potentially infinite" as their *explicantia*—such as "$\forall x(x \models T \Longrightarrow x$ is potentially infinite)" or "$\exists x(x \models T \wedge x$ is potentially infinite)".

Sadly, definitions with *definientia* containing "x is potentially infinite" are useless as long as the latter formula has not been explained satisfactorily. Now, there are model-theoretic *explicantia* of "T assumes (merely) the potential infinite" which avoid "a is potentially infinite" altogether. One may take "$\forall x$ ($x \models T \Longrightarrow x$ is finite)" or "$\exists x$ ($x \models T \wedge x$ is finite)".

These definitions make sense; but as we have already seen, they do not avoid trivializing counterexamples and, thus, cannot be accepted as explications. In comparison, employing (AF-T) or (CAF-T) as *explicantia* of "T assumes (merely) the potential infinite" is certainly more sophisticated, and it has a higher chance of being intuitively convincing. Yet, in this case, "T assumes (merely) the potential infinite" is not used as a means to motivate and explain the adoption of (AF-T) or (CAF-T)—it is just the other way round: the direction of the motivation is rather turned on its head. Nonetheless, taking (CAF-T) as an *explicans* of "T assumes (merely) the potential infinite" is just what we may end up with.

3. Conclusion

An upshot of the previous sections' considerations is: the proof theorist who does not subscribe to (D2) and does not want to fall back on Tait's analysis of "S is a finitistic theory" has to face the problem that he does not have a general explication of that phrase at his disposal. In this situation, he *may* well accept (D2) or (D2′).[60] For

[58] [24] may be read as suggesting a non-Tarskian semantics for schemas.

[59] Another idea could be to replace talk of objects by talk of, say, *processes*. In Mycielski's statement that a theory T "itself is regarded as a process of formation of axioms and theorems and not as an actually infinite object", this suggestion is clearly at work. Certainly, the stages of this process are finite objects or processes or … But what about T itself? It is infinite—actually infinite. In general, I think that if processes are that what is talked about, they are treated as objects; thus, I see no gain in such an approach.

[60] As I understand Lavine, he does assert that each theory T for which (CAF-T) holds belongs to finite mathematics; but it should be kept in mind that Lavine's assessment rests on a specific conception of theories.

he may dismiss the criticism of section 2.3 as being merely terminological. And the challenge implicit in the result of the Mycielski-translatability of ZF into a locally finite theory (i.e., into a theory belonging to finite mathematics in the sense of Lavine) need not bother him either: he may regard this relation as being too wide.

Let me begin with a discussion of the second reply. In fact, the assertion that each theorem of ZF has a *counterpart* in FIN(ZF) can be construed as a very weak claim. For example, if the n-th theorem of ZF is mapped to the n-th tautology "$(p \vee \neg p)_n$", we have it that each theorem of ZF has a counterpart even in propositional logic. It has to be noticed, though, that Lavine probably does intend his counterpart-relation to be not that trivial. Thus, in the abstract of [25], he tells us that "a system of finite mathematics is proposed that has all of the power of classical mathematics"[61]—which may remind us of Mycielski's assertion that FIN(ZF) is isomorphic to ZF.[62] Let me point out that neither [29, 30] nor [24, 25] contain a discussion of—let alone reasons for—the appropriateness of their terminologies.[63] In my opinion, employing the predicates "is isomorphic with" and "has the same power as" for what I have called "Mycielski-translation" (to have a "neutral" expression) is an exaggeration.[64]

What the proof theorist would consider to be disquieting should rather be a finitistic reduction of ZF. It can be shown, however, that

(2) ZF is proof-theoretically reducible to a theory T satisfying (CAF-T)

is the case. One way to obtain such a result is already contained in Mycielski's informal argument for the Mycielski-translatability of ZF into FIN(ZF); more generally, we have the

Observation.[65] For each recursively enumerable theory S, $S\ \overline{\rho_{\mathrm{un}}}\ \mathrm{FIN}(S)$ holds.

Sketch of the proof. An analysis of the argument for the corresponding informal claim shows that for each Mycielski-translation $\mathfrak{T}$,

$$\mathrm{Proof}_\sigma(x, y) \wedge \mathrm{ClEq}(y) \wedge lh(x) \leq z \longrightarrow \mathrm{Pr}_{\mathrm{FIN}(\sigma)}(\mathfrak{T}(y))$$

can be proved by a formalized induction on z.[66] Since this formula is Σ^0_1 and $\mathfrak{T}$ is the identity-function on atomic formulas, what can be shown is actually

$$\mathrm{I}\Sigma_1 \vdash \forall y\ (\mathrm{Pr}_\sigma(y) \wedge \mathrm{ClEq}(y) \longrightarrow \mathrm{Pr}_{\mathrm{FIN}(\sigma)}(y)).$$

This establishes the claim.[67]

[61] Lavine does not repeat this strong assertion.

[62] Of course, the trivial translation mentioned above does not show that ZF and propositional logic are *isomorphic* or *have the same power*.

[63] Moreover, there are no definitions of "is isomorphic with" and "has the same power as" in these texts.

[64] I do not argue for this claim here; but see [31] for more on this.

[65] In principle, this is noticed in [25].

[66] "FIN(σ)" is a Σ^0_0-formula of $\mathrm{L_{PA}}$ which represents the recursively enumerable set of axioms of FIN(S).

[67] This reasoning leaves it open whether $S\ \overline{\rho_{\mathrm{FIN[S]}}}\ \mathrm{FIN}(S)$ also holds.

(2) follows since (CAF-FIN(ZF)) is the case. Let me now present another argument for (2). In distinction to Mycielski's construction, it employs only general proof theoretical methods. I start with a simple, yet useful lemma.

Lemma 1. *Let L be an extension of L_{PRA}, T be a consistent theory in L, and let*

$$T\lceil := \{\psi \mid \psi \text{ is a formula of } L_{\mathrm{PRA}} \wedge T \vdash \psi\}.^{68}$$

If PRA $\subseteq T\lceil$, *then T is a conservative extension of $T\lceil$ (with respect to L_{PRA}) and* (CAF-$T\lceil$) *holds.*

Proof. The conservativity-claim is trivial. In order to show (CAF-$T\lceil$), let φ be a sentence of $\mathrm{L_{PRA}}$ such that $T\lceil \vdash \varphi$. Since φ is quantifier-free, either PRA $\vdash \varphi$ or PRA $\vdash \neg\varphi$; and because of PRA $\subseteq T\lceil$ and the consistency of $T\lceil$, PRA $\vdash \varphi$ has to be the case. Now, (CAF-PRA) holds; thus, φ has a finite model. □

Theorem 1. *Let S be a consistent recursively enumerable theory (in classical first-order logic); then there exist consistent recursively enumerable theories A, B such that*

(i) *S is relatively interpretable in B;*

(ii) *B is a conservative extension of A and* (CAF-A)*;*

(iii) *$B \overline{\rho_{\mathrm{un}}} A$;*

If, in addition, L_{S} is an extension of L_{PA} and $\mathrm{Q} \subseteq S$, *then*

(iv) *$S \overline{\rho_{\mathrm{un}}} B$ and $S \overline{\rho_{\mathrm{un}}} A$.*

Proof. For consistent axiomatizable S, let $B := (\mathrm{QF\text{-}IA}) + \mathrm{Con}_\sigma$ and let β, which is defined by

$$\beta(x) := (\mathrm{qf\text{-}ia})(x) \vee x = \overline{\ulcorner \mathrm{Con}_\sigma \urcorner},$$

be a Σ_0^0-formula of $\mathrm{L_{PA}}$ representing a set of axioms of B.

Let $A := B\lceil$, and set $\alpha(x) := \mathrm{Pr}_\beta(x) \wedge \mathrm{l[pra]}(x)$. α is a Σ_1^0-formula of $\mathrm{L_{PA}}$ representing a set of axioms of A, and Pr_α is a representation of $\overline{A}$ (formulated in $\mathrm{L_{PRA}}$).

(i) "S is relatively interpretable in B" follows from a theorem due to Feferman (see [3] and, for refinements, [47]).
(ii) Since $\mathrm{L_B}$ extends $\mathrm{L_{PRA}}$, Lemma 1 is applicable, and therefore (ii) is the case.
(iii) Because of the definition of α,

$$(\mathrm{QF\text{-}IA}) \vdash \forall x\, (\mathrm{Pr}_\beta(x) \wedge \mathrm{l[pra]}(x) \longrightarrow \mathrm{Pr}_\alpha(x))$$

[68] $T\lceil$ should be understood as being formulated in $\mathrm{L_{PRA}}$.

holds. Thus, we have

$$(\text{QF-IA}) \vdash \forall x\ (\text{Pr}_\beta(x) \wedge \text{ClEq}(x) \longrightarrow \text{Pr}_\alpha(x)),$$

from which (iii) follows by the Theorem cited in section 2.1.

(iv) By (QF-IA)-provable Σ^0_1-completeness of (qf-ia),

$$(\text{QF-IA}) \vdash \text{Pr}_\sigma(\overline{\ulcorner \bot \urcorner}) \longrightarrow \text{Pr}_{(\text{qf-ia})}(\overline{\ulcorner \text{Pr}_\sigma(\overline{\ulcorner \bot \urcorner}) \urcorner}),$$

whence

$$(\text{QF-IA}) \vdash \text{Con}_{(\text{qf-ia})+\text{Con}_\sigma} \longrightarrow \text{Con}_\sigma .$$

Due to the Theorem from section 2.1, this yields $S\ \overline{\rho_{\text{un}}}\ B$. $S\ \overline{\rho_{\text{un}}}\ A$ follows by the transitivity of $\overline{\rho_{\text{un}}}$ from the two previous claims. □

The moral of these theorems is, then, that if Definition 1, (D1) and (D2) are accepted, one has to swallow the fact that ZF is finitistically reducible.[69] And it is reducible to a potentially infinite theory, if (D2′) is chosen instead of (D2). As I said, in my opinion this is not only intuitively inacceptable; it would moreover be a highly unwelcome result for the existing reductive proof theory. For it would show the futility of the piece-meal attempts to proof-theoretically reduce larger and larger—yet, when compared to ZF, small—portions of mathematics to theories which are foundationally distinguished but, in general, non-finitary: all of those theories would even be proof-theoretically reducible to a—single—finitistic theory. But, of course, the above metatheorems do not imply that ZF *is* finitistically reducible (or reducible to a potentially infinite theory): for (D1) + (D2) (and (D1) + (D2′)), but also the reducibility relation employed, may be rejected as being inadequate to our preformal understanding of the corresponding concepts. Let me close this paper with some suggestions along these lines.

Dealing with the reducibility-concept first, it could be objected (similarly to the suggested rebuttal of Mycielski-translatability) that too many theories are relata of proof-theoretical reducibility. A natural proposal could be that nothing short of a relative interpretation, for example, should be accepted as a relation of inter-theoretical reduction.[70] In fact, using relative interpretability, we can generalize (P1). For this, let the theories considered be formulated in first-order languages, and let "$S \preceq T$" abbreviate "S is relatively interpretable in T":

Lemma 2.[71]

(i) *If $S \preceq T$ and all models of S are infinite, then all models of T are infinite.*

(ii) *If $S \preceq T$ and* (CAF-T)*, then* (CAF-S).

[69] I view L_{ZF} as an extension of L_{PA} in this case.

[70] See again [31, 32]; for a different opinion, see [7].

[71] This lemma also holds for local relative interpretability instead of (global) relative interpretability. See [29] for related results.

Proof.

(i) Assume that $\mathfrak{I}$ is a relative interpretation of S in T and $\forall \mathcal{M}$ ($\mathcal{M} \models S \Longrightarrow$ M is infinite). Then, because of the completeness theorem of first-order logic, for each $n \in \omega$, $S \vdash \exists_{\geq n}$. By the definition of "relative interpretation", $T \vdash \exists_{\geq n}$ follows for each $n \in \omega$, whence $\forall \mathcal{A}$ ($\mathcal{A} \models T \Longrightarrow A$ is infinite).

(ii) Assume that $S \preceq T$ and (CAF-T); then T and, therefore, S are consistent. Hence, if $\neg$(CAF$-S$), there is a sentence ψ in $\mathrm{L_S}$ such that

(∗) $S \vdash \psi$ and all models $\mathcal{M}$ of ψ have infinite domain

By the definition of "relative interpretation", there is a sentence φ in $\mathrm{L_T}$ such that

(∗∗) $$T \vdash \varphi \ \wedge \ \{\psi\} \preceq \{\varphi\}$$

Now, (i), (∗) and (∗∗) yield that $T \vdash \varphi$, but φ has only infinite models—contradicting (CAF-T). □

In the light of this lemma, the following modification of Definition 1 could be envisaged:

(D5) S is finitistically reducible : $\Longleftrightarrow$ there is a theory T such that (CAF-T) and $S \preceq T$

Whereas, given Definition 1 and (D1), (D2) was too wide, (D5) seems to be too narrow.[72] For example, by Lemma 2 and (D5), theories like (QF-IA) and Q are precluded from being finitistically reducible. In fact, Lemma 2 and (D5) imply that "S is finitistically reducible" is equivalent to (CAF-S). Furthermore, for theories T which are not formulated in a quantificational language—such as PRA in $\mathrm{L_{PRA}}$—"$S \preceq T$" and "$T \preceq S$" are not defined; thus, PRA would not be finitistically reducible for a trivial reason.[73]

[72] This negative assessment notwithstanding, I think that it is plausible that a reducibility relation ρ should satisfy at least a weakened analogue of (i) from Lemma 2: for S, T (which should perhaps again be recursively enumerable) formulated in first-order languages,

(P2): If $S \rho T$ and all models of S are infinite, then some model of T is infinite.

The intuition underlying (P2) is that a reducibility relation should respect the vast difference between the finite and the infinite and their respective assumptions. If S has only infinite models, it makes assumptions of infinity. If T has only finite models, then there exists a number n such that each of these models has less than n elements (for T is assumed to be a first-order theory); such a T avoids assumptions of infinity altogether. Even this cautious principle fails, however, if uniformly provable relative consistency is taken for ρ in (P2): let T be the first-order theory in $\mathrm{L_{PA}}$ with the single non-logical axiom "$\forall xy\ (x = y)$", and let τ be the *canonical representation* of this axiom-set, see [3]. Since (QF-IA) $\vdash \mathrm{Con_q}$, we also have (QF-IA) $\vdash \mathrm{Con_\tau} \rightarrow \mathrm{Con_q}$. But T has no infinite models.

[73] The usual definition of "S is relatively interpretable in T" can be quite naturally extended to theories S formulated in a quantifier-free fragment of a first-order language. Yet, there is at least one such extension which leads to the falsity of "PRA is finitistically reducible".

Surely, if (D1) and (D2) together with Definition 1 are maintained, we have no reasonable definition of "S is finitistically reducible"; and neither do we have one if these definitions are replaced by (D5). But even if (D1) + (D2) (and (D1) + (D2′)) are relinquished, the fact remains that each consistent recursively enumerable theory S is proof-theoretically reducible to a theory T such that (CAF-T) holds. Seen from this perspective, the problem is not so much that ZF is proof-theoretically reducible to a theory which has been called "finitistic" or "potentially infinitistic". It is rather that *being proof-theoretically reducible to a theory T such that* (*CAF-T*) *holds* is an empty criterion for distinguishing between recursively enumerable theories.[74] At this point, one could certainly start thinking about constructing reducibility concepts which conjoin the strengths of proof-theoretical reducibility and relative interpretability. However, what seems to be at stake *here* is rather to find a convincing explication of "T is a finitistic theory".[75] Now, the main problem is that apart from Tait's analysis, we merely seem to have (CAF-T) as a possible *explicans* of "T is a finitistic theory". But there are too many theories T for which (CAF-T) holds which nonetheless intuitively fail to be finitistic ones. If, however, (D1) + (D2) is rejected, which general reasons do we have to regard PRA as a finitary theory?[76] Thus, from my point of view, there remains a question at the end of this text: *What could a convincing general explication of "S is a finitistic theory" be?*[77]

Acknowledgement. This paper is an elaboration of a talk held at the 2001 Munich Russell Conference. I wish to thank the editor, Godehard Link, and Shaughan Lavine, Daniel Mook and Michael McLaughlin for useful comments.

References

[1] Aristotle: Physics. In: J. L. Ackrill (ed.), *A New Aristotle Reader*, Oxford: Clarendon Press, 1987.

[2] Drake, Frank R.: 1989. On the foundations of mathematics in 1987. In: H.-D. Ebbinghaus *et.al*. (eds.), *Logic Colloquium '87*, Amsterdam: North-Holland, 11–25.

[3] Feferman, Solomon: 1960. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae* XLIX: 35–92.

[74] This is also the answer to the claim that the criticism of section 2.3 is a purely terminological one.

[75] I think that from the components of the *definiens* of Definition 1, "S is reducible to T" is intuitively much clearer than "T is finitary".

[76] One could go on and ask: by which features is PRA distinguished in its role as a foundational theory? It is not only because (CAF-PRA) holds.

[77] In this text, I have presented this and related questions as being problems for the proof theorist rather than for Mycielski and Lavine. I have two reasons for this: First, in distinction to the proof theorist, Mycielski and Lavine should be quite content with (2); for them, the finitistic reducibility of ZF is simply no challenge. Second, with respect to this assessment my intuitions seem to be so far away from Mycielski's and Lavine's that I do not know what could count as an argument against (2) for them.

[4] Feferman, Solomon: 1988. Hilbert's program relativized: proof-theoretical and foundational reductions. *Journal of Symbolic Logic* 53: 364–384.

[5] Feferman, Solomon: 1991. Reflecting on incompleteness. *Journal of Symbolic Logic* 56: 1–49.

[6] Feferman, Solomon: 1993. What rests on what? The proof-theoretic analysis of mathematics. In: J. Czermak (ed.), *Philosophie der Mathematik*, Akten des 15. Internationalen Wittgenstein-Symposiums, Wien: Hölder-Pichler-Tempsky, 147–171.

[7] Feferman, Solomon: 2000. Does reductive proof theory have a viable rationale. *Erkenntnis* 53: 63–96.

[8] Girard, Jean-Yves: 1987. *Proof Theory and Logical Complexity*, vol. 1. Naples: Bibliopolis.

[9] Hájek, Petr and Pavel Pudlák: 1993. *Metamathematics of First-Order Arithmetic*. Berlin: Springer.

[10] Hilbert, David: 1923. Die logischen Grundlagen der Mathematik. *Mathematische Annalen* 88: 151–165. Reprinted in [16]: 178–191.

[11] Hilbert, David: 1926. Über das Unendliche. *Mathematische Annalen* 95: 161–190. Reprinted in [16]: 79–108. English translation 'On the infinite', in [46]: 367–392.

[12] Hilbert, David: 1928. Die Grundlagen der Mathematik, *Abhandlungen aus dem Mathematischen Seminar der Hamburger Universität* 6: 65–85 [english translation 'The foundations of mathematics', in [46]: 464–479].

[13] Hilbert, David: 1928. Probleme der Grundlegung der Mathematik. *Atti del Congresso nazionale dei matematici, Bologna 3–10 settembre 1928* 1: 135–141. With supplementations reprinted in Mathematische Annalen 102, 1929: 1–9.

[14] Hilbert, David: 1931. Die Grundlegung der elementaren Zahlenlehre. *Mathematische Annalen* 104: 485–494.

[15] Hilbert, David: 1931. Beweis des Tertium non datur. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, Mathematisch-physikalische Klasse: 120–125.

[16] Hilbert, David: 1964. *Hilbertiana*. Darmstadt: Wissenschaftliche Buchgesellschaft.

[17] Hilbert, David and Paul Bernays: 1968. *Grundlagen der Mathematik I*. Berlin: Springer.

[18] Hilbert, David and Paul Bernays: 1968. *Grundlagen der Mathematik II*. Berlin: Springer.

[19] Isaacson, Daniel: 1992. Some considerations on arithmetical truth and the ω-rule. In: M. Detlefsen (ed.), *Proof, Logic and Formalization*, London-New York: Routledge, 94–138.

[20] Kaye, Richard: 1991. *Models of Peano Arithmetic*. Oxford: Clarendon Press.

[21] Kreisel, Georg: 1958. Hilbert's programme. *Dialectica* 12: 346–372.

[22] Kreisel, Georg: 1960. Ordinal logics and the characterization of informal concepts of truth. *Proceedings of the International Congress of Mathematicians, Edinburgh 1958*, Cambridge: Cambridge University Press, 289–299.

[23] Kunen, Kenneth: 1980. *Set Theory*. Amsterdam: Elsevier.

[24] Lavine, Shaughan: 1994. *Understanding the Infinite*. Cambridge, MA: Harvard University Press.

[25] Lavine, Shaughan: 1995. Finite mathematics. *Synthese* 103: 389–420.

[26] Lorenzen, Paul: 1957. Das Aktual-Unendliche in der Mathematik. *Philosophia Naturalis* 4: 4–11.

[27] Martin, Richard M.: 1958. *Truth and Denotation*. London: Routledge & Kegan Paul.

[28] Mittelstraß, Jürgen (ed.): 1996. *Enzyklopädie Philosophie und Wissenschaftstheorie* 4. Stuttgart & Weimar: Metzler.

[29] Mycielski, Jan: 1981. Analysis without actual infinity. *Journal of Symbolic Logic* 46: 625–633.

[30] Mycielski, Jan: 1986. Locally finite theories. *Journal of Symbolic Logic* 51: 59–62.

[31] Niebergall, Karl-Georg: 2000. On the logic of reducibility: axioms and examples. *Erkenntnis* 53: 27–61.

[32] Niebergall, Karl-Georg: 2002. Structuralism, model theory and reduction. *Synthese* 130: 135–162.

[33] Niebergall, Karl-Georg and Matthias Schirn: 1998. Hilbert's finitism and the notion of infinity. In: M. Schirn (ed.), *The Philosophy of Mathematics today*, Oxford: Oxford University Press, 271–305.

[34] Niebergall, Karl-Georg and Matthias Schirn: 2002. Hilbert's programme and Gödel's theorems. *Dialectica* 56: 347–370.

[35] Niebergall, Karl-Georg and Matthias Schirn: 2003. What finitism could not be. *Critíca* 35: 43–68.

[36] Parsons, Charles: 1998. Finitism and intuitive knowledge. In: M. Schirn (ed.), *The Philosophy of Mathematics today*, Oxford: Oxford University Press, 249–270.

[37] Pudlák, Pavel: 1985. Cuts, consistency statements and interpretability. *Journal of Symbolic Logic* 50: 423–441.

[38] Rolf, Bertil: 1980. The finitary standpoint. *Erkenntnis* 15: 287–300.

[39] Schirn, Matthias and Karl-Georg Niebergall: 2001. Extensions of the finitist point of view. *History and Philosophy of Logic* 22: 135–161.

[40] Simpson, Stephen: 1988. Partial realizations of Hilbert's program. *Journal of Symbolic Logic* 53: 349–363.

[41] Simpson, Stephen: 1999. *Subsystems of Second Order Arithmetic*. Berlin: Springer.

[42] Smorynski, Craig: 1985. *Self-Reference and Modal Logic*. Berlin: Springer.

[43] Tarski, Alfred, Andrzej Mostowski, and Raphael M. Robinson: 1953. *Undecidable theories* Amsterdam: North-Holland.

[44] Tait, William W.: 1981. Finitism. *The Journal of Philosophy* 78: 524–546.

[45] Thomson, James: 1967 & 1981. Infinity in mathematics and logic. In: P. Edwards (ed.), *The Encyclopedia of Philosophy* vol. 4, New York: The Macmillan Company, 183–190.

[46] van Heijenoort, Jean (ed.): 1967. *From Frege to Gödel*. Cambridge, MA: Harvard University Press.

[47] Visser, Albert: 1991. The formalization of interpretability. *Studia Logica* 50: 81–105.

[48] Waismann, Friedrich: 1936. *Einführung in das mathematische Denken*. Wien: Gerold und Co.

Seminar für Philosophie, Logik und Wissenschaftstheorie
Philosophie-Department
Ludwig-Maximilians-Universität München
Ludwigstr. 31/I
80539 München
Germany

E-mail: kgn@lrz.uni-muenchen.de

Inconsistency in the Real World

Tobias Hürter

Abstract. It seems evident that there are infinitely many natural numbers—but can we be sure? To get a feeling for the strength of the this weakest axiom of infinity, it is useful to ask under what circumstances it may fail, while the elementary concept of number continues to hold. Here, we construct a hypothetical situation in which we are forced to conclude that a term for a large natural number cannot receive its intended interpretation. So from down below, we see that there is a bound on the natural numbers far up.

1. Number-Theoretic Science Fiction

"God created the natural numbers, everything else is man's work", goes Kronecker's famous *dictum*. But how many of them did He create? Certainly enough for everyday purposes. A spider has eight legs, and this eightness is a feature of the spider, independent of all mathematical theorizing. In this sense, natural numbers appear as given, almost like physical objects. However, physical theories describe the universe as finite and discrete: Inflationary cosmology gives an explicit estimate of the number of baryons in the universe from first principles: 10^{87}. The discreteness of space and time is particularly striking in the recent theory of Loop Quantum Gravity, see e.g., [4]. So it seems likely that numbers like $2^{2^{10}}$ have no physical manifestation. We have to derive formal properties of $2^{2^{10}}$ from our theories about the natural numbers, whereas we can tell that eight is even by direct acquaintance, or we can verify that $173 \cdot 12 = 2076$ by counting.[1] Edward Nelson distinguishes between a *genetic* and a *formal* notion of number:

> The point is that to regard $2 \Uparrow 5$ as standing for a genetic number entails a philosophical commitment to some ideal notion of existence. To a nominalist, $2 \Uparrow 4$ stands for a number, 65536, to which one can count, but $2 \Uparrow 5$ is a pair of arabic numerals with a double arrow between them, and there is not a scintilla of evidence that it stands for a genetic number. ([3]: 75)

[1] Can we verify by counting that 181 is prime? This is a borderline case, because we need the unrestricted monotonicity of multiplication, which involves a step of reflection on the process of counting.

On several occasions in the history of science, we saw our intuitions about apparently well-understood concepts break down outside the scales of our everyday world. Examples are the concepts of space, time and set. We cannot exclude the possibility that the concept of number will join this list.

To understand what goes on in the transition from the genetic to the formal notion of number, it is useful to ask under what circumstances the one holds while the other fails. The axioms of formal number theory (Peano arithmetic and its relatives) codify extrapolations from our pre-theoretical number concept, drawn from experiences in the real world. The commitment to infinitely many numbers or even of large finite numbers is the foremost of these extrapolations. It is the weakest axiom of infinity. Evident as it may appear, it is a substantial assumption. To investigate under what circumstances it could fail should be a worthwhile exercise—often it is healthy to step back and check your premises.

If the existence of a number is purely formal, we should be able to express the assertion of this existence in an adequate and consistent formal theory of numbers. But Gödel's second incompleteness theorem reminds us of the possibility that our usual axiom systems of number theory are inconsistent. The line behind which such an inconsistency would force us to retreat lies somewhere between the genetic and the formal notion. It would not change a spider's number of legs, but it would disrupt our view about the truth of some Π_1 sentences—and if things go really bad, it may force us to reconsider the existence of $2 \Uparrow 5$. Here I want to take a closer look at this worst-case scenario. Because we have no positive evidence for an inconsistency in Peano arithmetic, the argument has to be highly hypothetical.

At first sight, it is not clear whether such a situation can arise at all. It is conceivable that the genetic notion of number has no extension into the infinite, without any accessible evidence. I want to argue that this is not so: We can make sense of the idea that from down below, we see a bound for the natural numbers far up. For that, we start from a weak number theory T_0 intended to capture the strength (not necessarily the spirit) of the genetic concept. In particular, T_0 is indifferent to the existence of infinitely many numbers. We construct a Δ_0 formula $\varphi(x)$ and a number term t such that T_0 implies $\exists x\, \varphi(x)$, we cannot refute the consistency of $T_0 + \exists x < t\, \varphi(x)$ by elementary means, but $T_0 + \exists x < t\, \varphi(x)$ has models only up to a finite bound, because t cannot receive its standard interpretation. The key point of the construction is that we can arrange that for a given n, $\varphi(\bar{n})$ can be decided inside an initial segment well below the anomaly. φ is a bounded variant of the Gödel fixed point. My account relies heavily on [5].

2. The Theory T_0

Let L_A be the first-order language with non-logical symbols

$$<,\ +,\ \cdot,\ 0,\ 1.$$

This is the language of arithmetic. Let T_0 be the theory in L_A with the axioms for an irreflexive transitive order, the recursive definitions of addition and multiplication, i.e., the universal closures of

$$\neg x < x,$$
$$x < y \wedge y < z \rightarrow x < z,$$
$$x + 0 = x,$$
$$x + (y + 1) = (x + y) + 1,$$
$$x \cdot 0 = 0,$$
$$x \cdot (y + 1) = x \cdot y + x,$$

as well as the universal closures of

$$x + 1 = y + 1 \rightarrow x = y,$$
$$x < y \rightarrow x + 1 \leq y,$$
$$x < x + 1 \leftrightarrow x + 1 \neq 0,$$

and the Δ_0 induction scheme:

$$\varphi(0) \wedge \forall x\ (\varphi(x) \rightarrow \varphi(x + 1)) \rightarrow \forall x\, \varphi(x)$$

for each Δ_0 formula φ.

T_0 has arbitrarily large finite models, the smallest with domain $\{0\}$. It does not prove monotonicity of the arithmetical operations with regard to $<$. Adding

$$x + 1 \neq 0$$

to T_0 results in $I\Delta_0$.[2]

Lemma 1. *Let $\mathcal{M} = \langle M, <, +, \cdot, 0, 1\rangle$ be a model of T_0. Then $\mathcal{M}$ is finite iff $\langle M, +, 0\rangle$ is the cyclic group generated by* 1.

We define a canonical numeral $\bar{n}$ for each natural number n by recursion. The numeral for 0 is the symbol "0". If $n = 2k + 1$, let $\bar{n}$ be the string

$$\left((1 + 1) \cdot \bar{k} + 1\right).$$

If $n = 2k + 2$, let $\bar{n}$ be the string

$$\left((1 + 1) \cdot \bar{k} + 1 + 1\right).$$

[2]At first sight, it may seem contrived to allow 0 into the range of the successor operation. T_0 is not claimed to be a conceptually accurate codification of some pretheoretical notion of number. The choice of T_0 was rather a matter of mathematical taste. (However, as Godehard Link remarked, there may be some justification in cultural history for casting elementary counting in terms of remainder classes.) Alternatives would be to treat the arithmetical operations as ternary predicates representing possibly partial functions, or to allow for a largest number "many", which absorbs every nonzero companion argument in addition and multiplication. But caution: The three approaches are not equivalent.

The length of $\bar{n}$ grows in proportion to $\log_2(n)$.

We need a more efficient notation for large numbers of the form 2^{2^n}. Set

$$\begin{aligned} t_0 &\equiv (1+1)\,, \\ t_{n+1} &\equiv (t_n \cdot t_n)\,. \end{aligned}$$

So t_n represents 2^{2^n}. The length of t_n grows in proportion to $n \cdot 2^n$. Note that

$$\lim_{n\to\infty} \frac{n \cdot 2^n}{2^{2^n}} = 0.$$

Define sentences σ_n by

$$\begin{aligned} \sigma_0 &\equiv 1 < 1+1, \\ \sigma_{n+1} &\equiv \sigma_n \wedge \forall x\ (1 < x \leq t_n \to t_n < t_n \cdot x)\,. \end{aligned}$$

σ_n asserts that the universe has size at least $2^{2^n}+1$. The length of σ_n is bounded by (say) $30 \cdot (n+30)^2 \cdot 2^n$. Note that

$$\lim_{n\to\infty} \frac{30 \cdot (n+30)^2 \cdot 2^n}{2^{2^n}} = 0.$$

For every L_A-term $s(\vec{x})$ define a formula $\varepsilon_s(\vec{x})$ ("$s(\vec{x})$ exists") by induction on the complexity of s:

$$\begin{aligned} \varepsilon_0 \equiv \varepsilon_{x_i} &\equiv \top, \\ \varepsilon_1 &\equiv 0 < 1, \\ \varepsilon_{s_0+s_1} &\equiv \varepsilon_{s_0} \wedge \varepsilon_{s_1} \wedge (s_0 + s_1 \geq \max(s_0, s_1))\,, \\ \varepsilon_{s_0 \cdot s_1} &\equiv \varepsilon_{s_0} \wedge \varepsilon_{s_1} \wedge \forall y < s_1\,(s_0 \cdot (y+1) \geq s_0)\,. \end{aligned}$$

Note that

$$T_0 \vdash \varepsilon_{t_n} \leftrightarrow \sigma_n.$$

Now for every bounded L_A-formula $\varphi(\vec{x})$ define $\varphi^*(\vec{x})$ by another induction: For atomic φ, set

$$\begin{aligned} (t_0(\vec{x}) = t_1(\vec{x}))^* &\equiv \varepsilon_{t_0} \wedge \varepsilon_{t_1} \to t_0(\vec{x}) = t_1(\vec{x})\,, \\ (t_0(\vec{x}) < t_1(\vec{x}))^* &\equiv \varepsilon_{t_0} \wedge \varepsilon_{t_1} \to t_0(\vec{x}) < t_1(\vec{x})\,. \end{aligned}$$

For φ of higher complexity, we can assume that φ is given in negation-free prenex form, because over T_0, every formula has a canonical equivalent of this form. Set

$$\begin{aligned} (\varphi \wedge \psi)^* &\equiv \varphi^* \wedge \psi^*, \\ (\varphi \vee \psi)^* &\equiv \varphi^* \vee \psi^*, \end{aligned}$$

and

$$(\exists y < t(\vec{x})\, \varphi)^* \equiv \varepsilon_t \to \exists y < t(\vec{x})\, \varphi^*,$$
$$(\forall y < t(\vec{x})\, \varphi)^* \equiv \forall y < t(\vec{x})\, \varphi^*.$$

for $y \notin \{\vec{x}\}$. In a sufficiently rich metatheory, we have:

Lemma 2. *If $I\Delta_0$ proves a bounded formula φ, then T_0 proves φ^*.*

Idea of the proof. Let φ be given, as well as a model $\mathcal{M}$ of T_0. We show that $\mathcal{M}$ satisfies φ^*. We can assume that $\mathcal{M}$ has at least two elements. Expand L_A to L_A^M by adding constants for all elements of the domain M of $\mathcal{M}$. These constants have their canonical interpretation in $\mathcal{M}$. (We do not distinguish between the constants and their designations.) Unfold $\mathcal{M}$ by associating with every L_A^M-term a finite sequence of elements of M in the following way: $\langle a \rangle$ is assigned to every $a \in M$. For molecular t, treat the sequences as numerals with generalized digits and order them accordingly—so sequences of equal length are ordered lexicographically. For example, let t be $(t_0 + t_1)$ and assume that t_i was assigned $\langle \vec{a}^i \rangle$. To get the sequence for t, add $\langle \vec{a}^0 \rangle$ and $\langle \vec{a}^1 \rangle$ just like you learned to add decimal numerals in elementary school. If the resulting sequence has the form (say) $\langle b_0, b_1 \rangle$, then b_0 is $t^{\mathcal{M}}$, and b_1 counts how often you have to pass over $0^{\mathcal{M}}$ in the calculation of $\mathcal{M}$. Analogously, multiplication is performed by the elementary school algorithm.[3] This way, we arrive at a term model $\mathcal{M}'$ for L_A^M. Every term is interpreted by the associated sequence. $\mathcal{M}'$ models $I\Delta_0$. The injection

$$e : \mathcal{M} \to \mathcal{M}',\ a \mapsto \langle a \rangle$$

has the property

$$\mathcal{M}' \vdash \psi\left[e(\vec{a})\right] \Rightarrow \mathcal{M} \vdash \psi^*(\vec{a})$$

for all bounded $\psi(\vec{x})$ in L_A and all $\vec{a} \in \mathcal{M}$. ($\mathcal{M}$ is mapped onto an initial segment of $\mathcal{M}'$.) In particular, $\mathcal{M} \vdash \varphi$. □

Slightly less expensive, but much more tedious is the proof-theoretical route to Lemma 2 by inductively turning cut-free derivations in $I\Delta_0$ into derivations in T_0.

3. Metamathematics in T_0

We now want to arithmetize syntax in T_0. In general, this can be done adequately only up to a certain limit, because a finite model cannot have codes for all finite sequences. We proceed along the lines of Section V.3 of [2]. Many constructions done there for $I\Delta_0$ can be transferred to T_0 with the help of Lemma 2. With an eye on the things to

[3] This idea is taken from the proof of [2]: Theorem IV. 2. 2.

come, more elementary proofs are desirable. We will not carry out the details here, but only indicate some crucial points, aiming to push the limit of adequacy as far as possible. Define the Δ_0 formula $\pi\,(m, n, p)$ ("p codes the ordered pair (m, n)") by

$$\begin{aligned}\pi\,(m,n,p) \equiv\ & \exists x \le p\,(\varepsilon_{(1+1)\cdot x} \\ & \wedge([\varepsilon_{x\cdot(m+n)+m} \wedge (1+1)\cdot x = m+n+1 \wedge p = x\cdot(m+n)+m] \\ & \vee[\varepsilon_{x\cdot(m+n)+m+1} \wedge (1+1)\cdot x = m+n \wedge p = x\cdot(m+n+1)+m])).\end{aligned}$$

This provably satisfies the usual uniqueness conditions, but not totality of course.

With the ordered pair at hand, we apply Definition V.3.5 of [2] to get an efficient coding of finite sequences. An important property of this coding is that "s is a finite sequence", "m occurs in the sequence s" and "m is the i-th element of the sequence s" can be defined in T_0 by formulae with all quantifiers bounded by s. This suffices to develop syntax in T_0 as done for $I\Delta_0$ in [2]: V.3(g). By standard definitions, we represent terms, formulae, proofs, etc., as well as syntactical operations like substitution. In particular, we arrive at a Δ_0 formula $\mathrm{Bew}_\tau\,(x, y)$ for "x is a proof of y in the theory defined by τ". Here we take $\tau\,(x)$ to be a Δ_0 formula binumerating T_0 in infinite models of T_0. We can certainly demand that

$$T_0 \vdash \sigma_6 \wedge \varepsilon_{\bar{n}} \to \big(\tau\,(\bar{n}) \leftrightarrow \tau^*\,(\bar{n})\big)$$

for all n. The starred versions of the usual adequacy conditions can be derived in T_0.

For example, we can express the consistency of T_0 inside T_0 by

$$\neg\exists x\big(\mathrm{Bew}_\tau\,\big(x, \overline{\ulcorner 0 < 0 \urcorner}\big)\big)^*,$$

which by the way is a theorem of T_0. ($\ulcorner\theta\urcorner$ is the number which codes θ.)

4. "Armageddon"

Now let k be a natural number greater than eight. Apply the Fixed Point Lemma (for formulae with only y free) in $I\Delta_0$ to the formula

$$\sigma_k \to \neg\exists x < t_k\ \mathrm{Bew}_\tau\,(x, y)\,.$$

We arrive at a sentence φ_k equivalent over $I\Delta_0$ to

$$\sigma_k \to \neg\exists x < t_k\ \mathrm{Bew}_\tau\,\big(x, \overline{\ulcorner \varphi_k \urcorner}\big).$$

A closer look at our coding and the proof of the Fixed Point Lemma reveals that $\ulcorner\varphi_k\urcorner \ll 2^{2^k}$. Therefore

$$T_0 \vdash \sigma_k \to \varepsilon_{\overline{\ulcorner\varphi_k\urcorner}}.$$

From our remark on quantifier bounds in the coding of finite sequences together with our demand on τ it follows that

$$\varphi_k \Longleftrightarrow \varphi_k^*$$

over T_0. φ_k is bounded. By Lemma 2, φ_k is a fixed point over T_0. Informally, φ_k asserts that if the universe has a size of at least $2^{2^k} + 1$, there is no proof of φ_k below 2^{2^k}.

Lemma 3. *For every natural number $k > 8$, T_0 implies φ_k.*

Proof. Suppose not. Let $\mathcal{M}$ be a model of

$$T_0 + \neg\varphi_k.$$

Then the universe of $\mathcal{M}$ has a size of at least $2^{2^k} + 1$, and in $\mathcal{M}$, there is a proof from T_0 of φ_k with code $p < 2^{2^k}$. By correctness of the proof predicate, there is a natural number n such that p is the n-th successor of 0 in $\mathcal{M}$ and n codes a proof from T_0 of φ_k in the real world.[4] But $\mathcal{M}$ satisfies T_0, and therefore also φ_k. Contradiction. □

So T_0 implies φ_k. Can we hope to find a proof for φ_k in T_0? We better hope not! Any proof surveyable to human or computer surely must have a code below 2^{2^9}. (If you disagree, raise k.) Imagine we found a proof with code $p < 2^{2^k}$. Then p turns out to be a "witness for Armageddon" (Woodin's words).

Theorem 1. *Suppose $p < 2^{2^k}$ codes a proof of φ_k. Then T_0 has no models with a size greater than 2^{2^k}.*

Proof. The proof is that of Lemma 3 backwards. We show that T_0 implies $\neg\sigma_k$. Suppose that

$$\mathcal{M} \models T_0 + \sigma_k.$$

Then

$$\mathcal{M} \models \varepsilon_{t_k},$$

i.e., t_k has the standard interpretation. The same holds for $\bar{p}$ and $\overline{\ulcorner\varphi_k\urcorner}$. But now it follows that

$$\mathcal{M} \models \sigma_k \wedge \mathrm{Bew}_\tau\left(\bar{p}, \overline{\ulcorner\varphi_k\urcorner}\right),$$

which contradicts

$$\mathcal{M} \models \varphi_k.$$

□

So p forces us to conclude that 2^{2^k} does not exist.

[4] This is the crucial step from the formal to the genetic notion of number.

5. So What?

Let's look back at the nightmare vision we just made up. We constructed a hypothetical situation in which we see from down below (from the realm of the genetic numbers) that a term intended to designate a large natural number cannot receive its intended interpretation. Then the structure of arithmetic cannot be extended beyond a certain finite point. This would be very bad news for number theorists, but shepherds, accountants and astronomers wouldn't need to care. Because we just wanted to show that such a situation is *conceivable*, we were free to use metatheoretical infinity assumptions in our construction—as long as we keep them out of the object theory. So we laid out our scenario as a hypothetical *reductio ad absurdum* argument: We started from our received view of infinity. Then we imagined we find the dreaded p and are thrown back to T_0.[5] Note that instead of conveniently transferring known results from $I\Delta_0$ to T_0 by Lemma 2, we could do the work directly in T_0.

A natural reaction to this scenario is: "Well, it's just a roundabout argument that φ_k has no short proof in T_0." But this means taking the existence of infinitely many numbers for granted. If you seriously drop the presupposition of infinity, you cannot rule out the existence of a proof coded in p as above. Tomorrow, a bright mathematician might wake up and find one.

Acknowledgement. Thanks to Dieter Donder, Stefan Iwan, Godehard Link, Uwe Lück, Holger Sturm and especially Karl-Georg Niebergall for helpful comments and criticism.

References

[1] Hájek, Petr: 1984. On a new notion of partial conservativity. *Computation and Proof Theory* (Lecture Notes in Mathematics 1104). Berlin: Springer, 217–232.

[2] Hájek, Petr and Pavel Pudlák: 1993. *Metamathematics of First-Order Arithmetic*. Berlin: Springer.

[3] Nelson, Edward: 1986. *Predicative Arithmetic*. Princeton: Princeton University Press.

[4] Rovelli, Carlo and Lee Smolin: 1995. Discreteness of area and volume in quantum gravity. *Nuclear Physics* B442: 593–622.

[5] A modified construction could avoid the conceptual difficulties associated with the assumption of an *actual* inconsistency in formal number theory. For that, we could allow a nonstandard bound on the length of the inconsistency proof by working with versions of φ_k with k a parameter of the object theory. Then the parametrized version of Lemma 3 fails. If we accept the analogy, we could extrapolate back to the standard case. For example, we could investigate the possibility of local consistency proofs by excluding inconsistencies up to 2^k or 2^{2^k}, based on evidence up to a nonstandard k. Results of [1] warn of some pitfalls that may lure in this line of argument.

[5] Woodin, W. Hugh: 1998. The tower of Hanoi. In: H. G. Dales and G. Oliveri (eds.), *Truth in Mathematics*, Oxford: Oxford University Press, 329–351.

Technology Review
Helstorfer Straße 7
30625 Hannover
Germany

E-mail: tobias.huerter@technology-review.de

Predicativity, Circularity, and Anti-Foundation

Michael Rathjen

Abstract. The Anti-Foundation axiom, AFA, has turned out to be a versatile principle in set theory for modelling a plethora of circular and self-referential phenomena. This paper explores whether AFA and the most important tools emanating from it, such as the solution lemma and the co-recursion principle, can be developed on predicative grounds, that is to say, within a predicative theory of sets.

If one could show that most of the circular phenomena that have arisen in computer science do not require impredicative set existence axioms for their modelling, this would demonstrate that their circularity is of a different kind than the one which underlies impredicative definitions.

1. Introduction

Russell discovered his paradox in May of 1901 while working on his *Principles of Mathematics* [28]. In response to the paradox he developed his distinction of logical types. Although first introduced in [28], type theory found its mature expression five years later in his 1908 article *Mathematical Logic as Based on the Theory of Types* [29]. In [29] Russell commences with a list of contradictions to be solved. These include Epimenides' liar paradox, his own paradox, and Burali-Forti's antinomy of the greatest ordinal. In a first analysis he remarks: "In all the above contradictions (which are merely selections from an indefinite number) there is a common characteristic, which we may describe as self-reference or reflexiveness" ([29]: 224). On closer scrutiny he discerns the form of reflexiveness that is the common underlying root for the trouble as follows:

> Whatever we suppose to be the totality of propositions, statements about this totality generate new propositions which, on pain of contradiction, must lie outside the totality. It is useless to enlarge the totality, for that equally enlarges the scope of statements about the totality. ([29]: 224)

Here Russell declares very lucidly a ban on so-called *impredicative definitions*, first enunciated by Poincaré. An impredicative definition of an object refers to a presumed totality of which the object being defined is itself to be a member. For example, to define a set of natural numbers X as $X = \{n \in \mathbb{N} : \forall Y \subseteq \mathbb{N}\ F(n, Y)\}$ is impredicative since it involves the quantified variable 'Y' ranging over arbitrary subsets of the

natural numbers $\mathbb{N}$, of which the set X being defined is one member. Determining whether $\forall Y \subseteq \mathbb{N}\ F(n, Y)\}$ holds involves an apparent circle since we shall have to know in particular whether $F(n, X)$ holds—but that cannot be settled until X itself is determined. Impredicative set definitions permeate the fabric of Zermelo–Fraenkel set theory in the guise of the separation and replacement axioms as well as the powerset axiom.

The avoidance of impredicative definitions has also been called the *Vicious Circle Principle*. This principle was taken very seriously by Hermann Weyl:

> The deepest root of the trouble lies elsewhere: a field of possibilities open into infinity has been mistaken for a closed realm of things existing in themselves. As Brouwer pointed out, this is a fallacy, the Fall and Original Sin of set theory, even if no paradoxes result from it. ([34]: 243)

When it turned out that the ramified theory of types rendered much of elementary mathematics unworkable, Russell, oblivious of his original insight, introduced, as an ad hoc device and for entirely pragmatic reasons, his notorious *axiom of reducibility*. This 'axiom' says that every set of higher level is coextensive with one of lowest level. Thereby he reconstituted the abolished impredicative definitions through the back door, as it were, undermining the whole rationale behind a ramified hierarchy of types and levels to such an extent that it might just as well have been jettisoned altogether. Weyl derisively proclaimed:

> Russell in order to extricate himself from the affair, causes reason to commit hara-kiri, by postulating the above assertion in spite of its lack of support by any evidence ('axiom of reducibility'). In a little book *Das Kontinuum*, I have tried to draw the honest consequence and constructed a field of real numbers of the first level, within which the most important operations of analysis can be carried out. ([34]: 50)

In his analyzes of the contradictions and antinomies, Russell frequently equates the common culprit with self-reference (cf. above) and sees the problems as arising from *reflexive fallacies* ([29]: 230). He bans propositions of the form *"x is among x's"* ([28]: 105) and in his own so-called "no-class" theory of classes he explains that propositional functions, such as the function *"x is a set"*, may not be applied to themselves as self-application would give rise to a vicious circle. Unlike his cogent analysis of the problem of impredicativity, Russell's charges against self-referential notions are not that well explained. It might be that the scope of his criticism was just intended to be confined to questions of concern to his own type theories, though it is more likely that he regarded self-referential notions as being incoherent or rather fallacious conceptions that should be banned all the same. In the history of philosophy, the charge of circularity is old and regarded as tantamount to delivering a refutation. In his *Analytica Posteriora*, Aristotle outlaws circular lines of arguments. Similarly, Aquinas calls an infinite series of reasons each of which is in some sense dependent on a prior a *vicious regress*. On the other hand, in hermeneutics it has been a tenet that comprehension can only come about through a tacit foreknowledge, thereby emphasizing the

inherent circularity of all understanding (the *hermeneutic circle*). In the same vein, circularity has been found in much intentional activity related to self-consciousness, communication, common knowledge, and conventions. For instance, on D. Lewis's account in his book *Convention* [20], for something to be a convention among a group of people, common knowledge about certain facts must obtain in this group. If C and E are modal operators such that $C\varphi$ and $E\varphi$ stand for *'φ is common knowledge'* and *'φ is known to everybody'*, respectively, then the central axiom which implicitly defines C takes the self-referential form $C\varphi \leftrightarrow E(\varphi \wedge C\varphi)$.

In logic, circular concepts involving self-reference or self-application have proven to be very important, as witnessed by Gödel's incompleteness theorems, the recursion theorem in recursion theory, self-applicative programs, and such so-called applicative theories as Feferman's *Explicit Mathematics* that have built-in gadgets allowing for self-application.

The general blame that Russell laid on circularity was more influential then the more specific ban placed on impredicative definitions. In the wake of Russell's anathematizing of circularity, Tarski's hierarchical approach via meta-languages became the accepted wisdom on the semantical paradoxes. On the other hand, the fruitfulness of circular notions in many areas of scientific discourse demonstrated that these notions are scientifically important. Thus one is naturally led to search for criteria that would enable one to tell the benign (and fruitful) circularities from the paradoxical ones. Kripke, though, in his *Outline of a Theory of Truth* [19], demonstrated rather convincingly that there are no such general 'syntactic' criteria for making this distinction, in that whether or not something is paradoxical may well depend on non-linguistic facts.

Notwithstanding the lack of simple syntactical criteria for detecting paradoxical circularities, it is of great interest to develop frameworks in which many non-paradoxical circular phenomena can be modelled. One such framework is Feferman's theory of explicit mathematics, cf. [13]. It is suitable for representing Bishop-style constructive mathematics as well as generalized recursion, including direct expression of structural concepts which admit self-application. Another very systematic toolbox for building models of various circular phenomena is set theory with the *Anti-Foundation axiom*. Theories like ZF outlaw sets like $\Omega = \{\Omega\}$ and infinite chains of the form $\Omega_{i+1} \in \Omega_i$ for all $i \in \omega$ on account of the Foundation axiom, and sometimes one hears the mistaken opinion that the only coherent conception of sets precludes such sets. The fundamental distinction between well-founded and non-well-founded sets was formulated by Mirimanoff in 1917. The relative independence of the Foundation axiom from the other axioms of Zermelo–Fraenkel set theory was announced by Bernays in 1941 but did not appear until the 1950s. Versions of axioms asserting the existence of non-well-founded sets were proposed by Finsler (1926). The ideas of Bernays' independence proof were exploited by Rieger, Hájek, Boffa, and Felgner. After Finsler, Scott in 1960 appears to have been the first person to consider an Anti-Foundation axiom which encapsulates a strengthening of the axiom of extensionality. The Anti-Foundation axiom in its strongest version was first formulated by Forti and Honsell [16] in 1983. Though several logicians explored set theories whose universes contained non-wellfounded sets

(or hypersets as they are called nowadays) the area was considered rather exotic until these theories were put to use in developing rigorous accounts of circular notions in computer science, cf. [3]. It turned out that the *Anti-Foundation Axiom,* AFA, gave rise to a rich universe of sets and provided an elegant tool for modelling all sorts of circular phenomena. The application areas range from modal logic, knowledge representation and theoretical economics to the semantics of natural language and programming languages. The subject of hypersets and their applications is thoroughly developed in the books [3] by P. Aczel and [5] by J. Barwise and L. Moss.

While reading [3] and [5], the question that arose in my mind was that of whether or not the material could be developed on the basis of a constructive universe of hypersets rather than a classical and impredicative one. This paper explores whether AFA and the most important tools emanating from it, such as the solution lemma and the corecursion principle, can be developed on predicative grounds, that is to say, within a predicative theory of sets. The upshot will be that most of the circular phenomena that have arisen in computer science don't require impredicative set existence axioms for their modelling, thereby showing that their circularity is clearly of a different kind than the one which underlies impredicative definitions.

2. The Anti-Foundation Axiom

Definition 2.1. A *graph* will consist of a set of *nodes* and a set of *edges*, each edge being an ordered pair $\langle x, y\rangle$ of nodes. If $\langle x, y\rangle$ is an edge then we will write $x \to y$ and say that y is a *child* of x.

A *path* is a finite or infinite sequence $x_0 \to x_1 \to x_2 \to \cdots$ of nodes $x_0, x_1, x_2, \ldots$ linked by edges $\langle x_0, x_1\rangle, \langle x_1, x_2\rangle, \ldots$.

A *pointed graph* is a graph together with a distinguished node x_0 called its *point*. A pointed graph is *accessible* if for every node x there is a path $x_0 \to x_1 \to x_2 \to \cdots \to x$ from the point x_0 to x.

A *decoration* of a graph is an assignment d of a set to each node of the graph in such a way that the elements of the set assigned to a node are the sets assigned to the children of that node, i.e.,

$$d(a) = \{d(x) : \ a \to x\}.$$

A *picture* of a set is an accessible pointed graph (apg for short) which has a decoration in which the set is assigned to the point.

Definition 2.2. The *Anti-Foundation Axiom*, AFA, is the statement that every graph has a unique decoration.

Note that AFA has the consequence that every apg is a picture of a unique set.

AFA is in effect the conjunction of two statements:

- AFA$_1$: *Every graph has at least one decoration.*
- AFA$_2$: *Every graph has a most one decoration.*

AFA$_1$ is an existence statement whereas AFA$_2$ is a strengthening of the Extensionality axiom of set theory. For example, taking the graph $\mathbb{G}_0$ to consist of a single node x_0 and one edge $x_0 \to x_0$, AFA$_1$ ensures that this graph has a decoration $d_0(x) = \{d_0(y) : x \to y\} = \{d_0(x)\}$, giving rise to a set b such that $b = \{b\}$. However, if there is another set c satisfying $c = \{c\}$, the Extensionality axiom does not force b to be equal to c, while AFA$_2$ yields $b = c$. Thus, by AFA there is exactly one set Ω such that $\Omega = \{\Omega\}$.

Another example which demonstrates the extensionalizing effect of AFA$_2$ is provided by the graph $\mathbb{G}_\infty$ which consists of the infinitely many nodes x_i and the edges $x_i \to x_{i+1}$ for each $i \in \omega$. According to AFA$_1$, $\mathbb{G}_\infty$ has a decoration. As $d_\infty(x_i) = \Omega$ defines such a decoration, AFA$_2$ entails that this is the only one, whereby the different graphs $\mathbb{G}_0$ and $\mathbb{G}_\infty$ give rise to the same non-well-founded set.

The most important applications of AFA arise in connection with solving systems of equations of sets. In a nutshell, this is demonstrated by the following example. Let p and q be arbitrary fixed sets. Suppose we need sets x, y, z such that

$$\begin{aligned} x &= \{x, y\} \\ y &= \{p, q, y, z\} \\ z &= \{p, x, y\}. \end{aligned} \tag{1}$$

Here p and q are best viewed as atoms while x, y, z are the indeterminates of the system. AFA ensures that the system (1) has a unique solution. There is a powerful technique that can be used to show that systems of equations of a certain type have always unique solutions. In the terminology of [5] this is called the *solution lemma*. We shall prove it in the sections on applications of AFA.

3. AFA in Constructive Set Theory

In this section I will present some results about the proof-theoretic strength of systems of constructive set theory with AFA instead of $\in$-Induction.

3.1. Constructive Set Theory

Constructive set theory grew out of Myhill's [23] endeavours to discover a simple formalism that relates to Bishop's constructive mathematics as ZFC relates to classical Cantorian mathematics. Later on Aczel modified Myhill's set theory to a system which he called Constructive Zermelo–Fraenkel set theory, CZF, and corroborated its

constructiveness by interpreting it in Martin-Löf type theory (MLTT), cf. [1]. The interpretation was in many ways canonical and can be seen as providing CZF with a standard model in type theory.

Let CZF^- be CZF without $\in$-induction and let CZFA be CZF^- plus AFA. I. Lindström [21] showed that CZFA can be interpreted in MLTT as well. Among other sources, the work of [21] will be utilized in calibrating the exact strength of various extensions of CZFA, in particular ones with inaccessible set axioms. The upshot is that AFA does not yield any extra proof-theoretic strength on the basis of constructive set theory and is indeed much weaker in proof strength than $\in$-Induction. This contrasts with Kripke–Platek set theory, KP. The theory KPA, which adopts AFA in place of the Foundation Axiom scheme, is proof-theoretically considerably stronger than KP as was shown in [26]. On the other hand, while being weaker in proof-theoretic strength, CZFA seems to be "mathematically" stronger than KPA in that most applications that AFA has found can be easily formalized in CZFA whereas there are serious difficulties with doing this in KPA. For instance, the proof from AFA that the collection of streams over a given set A exists and forms a set, seems to require the exponentiation axiom, a tool which is clearly not available in KPA.

3.2. The Theory CZFA

The language of CZF is the first order language of Zermelo–Fraenkel set theory, LST, with the non logical primitive symbol $\in$. We assume that LST has also a constant, ω, for the set of the natural numbers.

Definition 3.1 (Axioms of CZF). CZF is based on intuitionistic predicate logic with equality. The set theoretic axioms of CZF are the following:

1. *Extensionality.* $\forall a\, \forall b\, (\forall y\, (y \in a \leftrightarrow y \in b) \rightarrow a = b)$.

2. *Pair.* $\forall a\, \forall b\, \exists x\, \forall y\, (y \in x \leftrightarrow y = a \vee y = b)$.

3. *Union.* $\forall a\, \exists x\, \forall y\, (y \in x \leftrightarrow \exists z{\in}a\; y \in z)$.

4. Δ_0*-Separation scheme.* $\forall a\, \exists x\, \forall y\, (y \in x \leftrightarrow y \in a \wedge \varphi(y))$,

   for every *bounded* formula $\varphi(y)$, where a formula $\varphi(x)$ is bounded, or Δ_0, if all the quantifiers occurring in it are bounded, i.e., of the form $\forall x{\in}b$ or $\exists x{\in}b$.

5. *Subset Collection scheme.*

$$\forall a\, \forall b\, \exists c\, \forall u\, \big(\forall x{\in}a\, \exists y{\in}b\; \varphi(x, y, u) \rightarrow \exists d{\in}c\; (\forall x{\in}a\, \exists y{\in}d\; \varphi(x, y, u) \wedge \forall y{\in}d\, \exists x{\in}a\; \varphi(x, y, u))\big)$$

   for every formula $\varphi(x, y, u)$.

6. *Strong Collection scheme.*

$$\forall a \left(\forall x{\in}a\, \exists y\, \varphi(x, y) \rightarrow \exists b\, (\forall x{\in}a\, \exists y{\in}b\, \varphi(x, y) \wedge \forall y{\in}b\, \exists x{\in}a\, \varphi(x, y))\right)$$

for every formula $\varphi(x, y)$.

7. *Infinity.*

$$(\omega 1) \qquad 0{\in}\omega \wedge \forall y\, (y{\in}\omega \rightarrow y + 1{\in}\omega)$$
$$(\omega 2)\ \forall x\, (0{\in}x \wedge \forall y\, (y{\in}x \rightarrow y + 1{\in}x) \rightarrow \omega{\subseteq}x),$$

where $y + 1$ is $y \cup \{y\}$, and 0 is the empty set, defined in the obvious way.

8. $\in$*-Induction scheme.*

$$(IND_{\in}) \qquad \forall a\, (\forall x{\in}a\, \varphi(x) \rightarrow \varphi(a)) \rightarrow \forall a\, \varphi(a),$$

for every formula $\varphi(a)$.

Definition 3.2. Let CZF$^-$ be the system CZF without the $\in$ - Induction scheme.

Remark 3.3. CZF$^-$ is strong enough to show the existence of any primitive recursive function on ω and therefore Heyting arithmetic can be interpreted in CZF$^-$ in the obvious way. By way of example, let's verify this for addition: As a consequence of Subset Collection one obtains that for arbitrary sets a, b, the class of all functions from a to b, ${}^a b$, is a set. Using Strong Collection, $\{{}^n\omega : n{\in}\omega\}$ is a set, and thus $\mathrm{Fin} := \bigcup_{n\in\omega} {}^n\omega$ is a set, too. Employing the axiom $(\omega 2)$ one shows that

$$\forall \langle n, m\rangle \in \omega \times \omega\, \exists! f \in \mathrm{Fin}\, \theta(n, m, f),$$

where $\theta(n, m, f)$ stands for the formula

$$\mathrm{dom}(f) = m + 1 \ \wedge\ f(0) = n \ \wedge\ (\forall i{\in}m) f(i + 1) = f(i) + 1.$$

Using Strong Collection, there exists a set A such that

$$\forall \langle n, m\rangle \in \omega \times \omega\, \exists f \in A\, \theta(n, m, f) \ \wedge\ \forall f{\in}A\, \exists \langle n, m\rangle \in \omega \times \omega\, \theta(n, m, f).$$

Now define $h : \omega \times \omega \rightarrow \omega$ by letting $h(n, m) = k$ if and only if

$$\exists f \in A \left[\theta(n, m, f) \ \wedge\ f(m) = k\right].$$

It is easy to show that h satisfies the recursion equations

$$h(n, 0) = n \quad \wedge \quad h(n, m + 1) = h(n, m) + 1.$$

Definition 3.4. Unfortunately, CZF$^-$ has certain defects from a mathematical point of view in that this theory appears to be too limited for proving the existence of the

transitive closure of an arbitrary set. To remedy this we shall consider an axiom, TRANS, which ensures that every set is contained in a transitive set:

$$\text{TRANS} \qquad \forall x\, \exists y \left[x \subseteq y \ \wedge\ (\forall u{\in}y)\, (\forall v{\in}u)\, v{\in}y \right].$$

Let CZFA be the theory $\mathrm{CZF}^- + \mathrm{TRANS} + \mathrm{AFA}$.

Lemma 3.5. *Let* $\mathrm{TC}(x)$ *stand for the smallest transitive set that contains all elements of* x. $\mathrm{CZF}^- + \mathrm{TRANS}$ *proves the existence of* $\mathrm{TC}(x)$ *for any set* x.

Proof. We shall use a consequence of Subset Collection called *Exponentiation* which asserts that for arbitrary sets a, b, the class of all functions from a to b, ${}^a b$, is a set.

Let x be an arbitrary set. By TRANS there exists a transitive set A such that $x \subseteq A$. For $n{\in}\omega$ let

$$\begin{aligned} B_n &= \{f \in {}^{n+1}A : f(0) \in x \ \wedge\ (\forall i{\in}n)\, f(i+1) \in f(i)\}, \\ \mathrm{TC}_n(x) &= \bigcup \{\mathrm{ran}(f) : f \in B_n\}, \end{aligned}$$

where $\mathrm{ran}(f)$ denotes the range of a function f. B_n is a set owing to Exponentiation and Δ_0 Separation. $\mathrm{TC}_n(x)$ is a set by Union. Furthermore, $C = \bigcup_{n\in\omega} \mathrm{TC}_n(x)$ is a set by Strong Collection and Union. Then $x = \mathrm{TC}_0(x) \subseteq C$. Let y be a transitive set such that $x \subseteq y$. By induction on n one easily verifies that $\mathrm{TC}_n(x) \subseteq y$, and hence $C \subseteq y$. Moreover, C is transitive. Thus C is the smallest transitive set which contains all elements of x. □

Definition 3.6. A mathematically very useful axiom to have in set theory is the *Dependent Choices Axiom*, DC, i.e., for all formulae ψ, whenever

$$(\forall x{\in}a)\, (\exists y{\in}a)\, \psi(x, y)$$

and $b_0{\in}a$, then there exists a function $f : \omega \to a$ such that $f(0) = b_0$ and

$$(\forall n{\in}\omega)\, \psi(f(n), f(n+1)).$$

For a function f let $\mathrm{dom}(f)$ denote the domain of f. Even more useful in constructive set theory is the *Relativized Dependent Choices Axiom*, RDC.[1] It asserts that for arbitrary formulae ϕ and ψ, whenever

$$\forall x \big[\phi(x) \ \to\ \exists y \big(\phi(y) \ \wedge\ \psi(x, y) \big) \big]$$

and $\phi(b_0)$, then there exists a function f with $\mathrm{dom}(f) = \omega$ such that $f(0) = b_0$ and

$$(\forall n{\in}\omega) \big[\phi(f(n)) \ \wedge\ \psi(f(n), f(n+1)) \big].$$

A restricted form of RDC where ϕ and ψ are required to be Δ_0 formulas will be called Δ_0-RDC.

[1] In [2], RDC is called the dependent choices axiom and DC is dubbed the axiom of limited dependent choices. We deviate from the notation in [2] as it deviates from the usage in classical set theory texts.

The defect of CZF^- concerning the lack of enough transitive sets can also be remedied by adding Δ_0-RDC to CZF^-. It is perhaps worth noting that Δ_0-RDC implies DC on the basis of CZF^-.

Lemma 3.7. $\mathrm{CZF}^- + \Delta_0\text{-RDC} \vdash \mathrm{DC}$.

Proof. See [27]: Lemma 3.4. □

The existence of the transitive closure of any set can also be obtained by slightly strengthening induction on ω to

$$\Sigma_1\text{-IND}_\omega \;\; \phi(0) \;\wedge\; (\forall n{\in}\omega)(\phi(n) \to \phi(n+1)) \;\to\; (\forall n{\in}\omega)\phi(n)$$

for all Σ_1 formulae ϕ. It is worth noting that $\Sigma_1\text{-IND}_\omega$ actually implies

$$\Sigma\text{-IND}_\omega \;\; \theta(0) \;\wedge\; (\forall n{\in}\omega)(\theta(n) \to \theta(n+1)) \;\to\; (\forall n{\in}\omega)\theta(n)$$

for all Σ formulae θ, where the Σ formulae are the smallest collection of formulae comprising the Δ_0 formulae which is closed under $\wedge$, $\vee$, bounded quantification, and (unbounded) existential quantification. This is due to the fact that every Σ formula is equivalent to a Σ_1 formula provably in CZF^-. The latter principle is sometimes called the Σ *Reflection Principle* and can be proved as in Kripke–Platek set theory (one easily verifies that the proof of ([4]: I.4.3) also works in CZF^-). $\Sigma\text{-IND}_\omega$ enables one to introduce functions by Σ recursion on ω ([4]: I.6) as well as the transitive closure of an arbitrary set (on the basis of CZF^-). It is worth noting that $\Sigma\text{-IND}_\omega$ is actually a consequence of Δ_0-RDC.

Lemma 3.8. $\mathrm{CZF}^- + \Delta_0\text{-RDC} \vdash \Sigma\text{-IND}_\omega$.

Proof. Suppose $\theta(0) \wedge (\forall n{\in}\omega)(\theta(n) \to \theta(n+1))$, where $\theta(n)$ is of the form $\exists x \phi(n, x)$ with ϕ Δ_0. We wish to prove $(\forall n{\in}\omega)\theta(n)$.

If z is an ordered pair $\langle x, y\rangle$ let $1^{\mathrm{st}}(z)$ denote x and $2^{\mathrm{nd}}(z)$ denote y. Since $\theta(0)$ there exists a set x_0 such that $\phi(0, x_0)$. Put $a_0 = \langle 0, x_0\rangle$.

From $(\forall n{\in}\omega)(\theta(n) \to \theta(n+1))$ we can conclude

$$(\forall n{\in}\omega)\forall y\,[\phi(n, y) \to \exists w\, \phi(n+1, w)\,]$$

and thus

$$\forall z\,\big[\,\psi(z) \to \exists v\,(\,\psi(v) \;\wedge\; \chi(z, v)\,)\big],$$

where $\psi(z)$ stands for *z is an ordered pair* $\wedge$ $1^{\mathrm{st}}(z) \in \omega$ $\wedge$ $\phi(1^{\mathrm{st}}(z), 2^{\mathrm{nd}}(z))$ and $\chi(z, v)$ stands for $1^{\mathrm{st}}(v) = 1^{\mathrm{st}}(z) + 1$. Note that ψ and χ are Δ_0. We also have $\psi(a_0)$. Thus by Δ_0-RDC there exists a function $f : \omega \to V$ such that $f(0) = a_0$ and

$$(\forall n{\in}\omega)\,\big[\,\psi(f(n)) \;\wedge\; \chi(f(n), f(n+1))\,\big].$$

From $\chi(f(n), f(n+1))$, using induction on ω, one easily deduces that $1^{\mathrm{st}}(f(n)) = n$ for all $n{\in}\omega$. Hence from $(\forall n{\in}\omega)\,\psi(f(n))$ we get $(\forall n{\in}\omega)\,\exists x\,\phi(n, x)$ and so $(\forall n{\in}\omega)\,\theta(n)$. □

We shall consider also the full scheme of induction on ω,

$$\mathrm{IND}_\omega \;\; \psi(0) \;\wedge\; (\forall n{\in}\omega)(\psi(n) \to \psi(n+1)) \;\to\; (\forall n{\in}\omega)\psi(n)$$

for all formulae ψ.

Lemma 3.9. $\mathrm{CZF}^- + \mathrm{RDC} \vdash \mathrm{IND}_\omega$.

Proof. Suppose $\theta(0) \wedge (\forall n{\in}\omega)(\theta(n) \to \theta(n+1))$. We wish to prove $(\forall n{\in}\omega)\theta(n)$. Let $\phi(x)$ and $\psi(x, y)$ be the formulas $x{\in}\omega \wedge \theta(x)$ and $y = x+1$, respectively. Then $\forall x\,[\phi(x) \to \exists y\,(\phi(y) \wedge \psi(x, y))]$ and $\phi(0)$. Hence, by RDC, there exists a function f with domain ω such that $f(0) = 0$ and $\forall n{\in}\omega\,[\phi(f(n)) \wedge \psi(f(n), f(n+1))]$. Let $a = \{n{\in}\omega :\; f(n) = n\}$. Using the induction principle ($\omega 2$) one easily verifies $\omega \subseteq a$, and hence $f(n) = n$ for all $n{\in}\omega$. Hence, $\phi(n)$ for all $n{\in}\omega$, and thus $(\forall n{\in}\omega)\theta(n)$. □

4. Predicativism

Weyl rejected the platonist philosophy of mathematics as manifested in impredicative existence principles of Zermelo–Fraenkel set theory. In his book *Das Kontinuum*, he initiated a predicative approach to the real numbers and gave a viable account of a substantial chunk of analysis. What are the ideas and principles upon which his "predicative view" is supposed to be based? A central tenet is that there is a fundamental difference between our understanding of the concept of natural numbers and our understanding of the set concept. Like the French predicativists, Weyl accepts the completed infinite system of natural numbers as a point of departure. He also accepts classical logic but just works with sets that are of level one in Russell's ramified hierarchy, in other words only with the principle of arithmetical definitions.

Logicians such as Wang, Lorenzen, Schütte, and Feferman then proposed a foundation of mathematics using layered formalisms based on the idea of predicativity which ventured into higher levels of the ramified hierarchy. The idea of an autonomous progression of theories $RA_0, RA_1, \ldots, RA_\alpha, \ldots$ was first presented in Kreisel's [18] and than taken up by Feferman and Schütte to determine the limits of predicativity. The notion of autonomy therein is based on introspection and should perhaps be viewed as a 'boot-strap' condition. One takes the structure of natural numbers as one's point of departure and then explores through a process of active reflection what is implicit in accepting this structure, thereby developing a growing body of ever higher layers of the ramified hierarchy. Feferman and Schütte [30, 31, 11, 12] showed that the ordinal Γ_0 is the first ordinal whose well-foundedness cannot be proved in autonomous progressions of theories. It was also argued by Feferman that the whole sequence of autonomous progressions of theories is coextensive with predicativity, and on these grounds Γ_0 is often referred to as the proper limit of all predicatively provable ordinals. In this paper I shall only employ the "lower bound" part of this analysis, i.e., that

every ordinal less than Γ_0 is a predicatively provable ordinal. In consequence, every theory with proof-theoretic ordinal less than Γ_0 has a predicative consistency proof and is moreover conservative over a theory RA_α for arithmetical statements for some $\alpha < \Gamma_0$. As a shorthand for the above I shall say that a theory is *predicatively justifiable*. The remainder of this section lists results showing that CZFA and its variants are indeed predicatively justifiable.

As a scale for measuring the proof-theoretic strength of theories one uses traditionally certain subsystems of second order arithmetic, see [14, 33]. Relevant to the present context are systems based on the Σ^1_1 axiom of choice and the Σ^1_1 axiom of dependent choices. The theory Σ^1_1-AC is a subsystem of second order arithmetic with the Σ^1_1 axiom of choice and induction over the natural numbers for all formulas while Σ^1_1-DC$_0$ is a subsystem of second order arithmetic with the Σ^1_1 axiom of dependent choices and induction over the natural numbers restricted to formulas without second order quantifiers (for precise definitions see [14, 33]). The proof theoretic ordinal of Σ^1_1-AC is $\varphi\varepsilon_0 0$ while Σ^1_1-DC$_0$ has the smaller proof-theoretic ordinal $\varphi\omega 0$ as was shown by Cantini [7]. Here φ denotes the Veblen function, see [32].

Theorem 4.1. (i) *The theories* CZF$^-$ + Σ_1-IND$_\omega$, CZFA + Σ_1IND$_\omega$ + Δ_0-RDC, CZFA + Σ_1-IND$_\omega$ + DC, *and* Σ^1_1-DC$_0$ *are proof-theoretically equivalent. Their proof-theoretic ordinal is* $\varphi\omega 0$.

(ii) *The theories* CZF$^-$ + IND$_\omega$, CZFA + IND$_\omega$ + RDC, $\widehat{\mathrm{ID}}_1$, *and* Σ^1_1-AC *are proof-theoretically equivalent. Their proof-theoretic ordinal is* $\varphi\varepsilon_0 0$.

(iii) CZFA *has at least proof-theoretic strength of Peano arithmetic and so its proof-theoretic ordinal is at least* ε_0. *An upper bound for the proof-theoretic ordinal of* CZFA *is* $\varphi 20$. *In consequence,* CZFA *is proof-theoretically weaker than* CZFA + Δ_0-RDC.

(iv) *All the foregoing theories are predicatively justifiable.*

Proof. (ii) follows from [27]: Theorem 3.15.

As to (i) it is important to notice that the scheme dubbed Δ_0-RDC in [27] is not the same as Δ_0-RDC in the present paper. In [27], Δ_0-RDC asserts for Δ_0 formulas ϕ and ψ that whenever $(\forall x{\in}a)\big[\phi(x) \rightarrow (\exists y{\in}a)\big(\phi(y) \wedge \psi(x,y)\big)\big]$ and $b_0{\in}a \wedge \phi(b_0)$, then there exists a function $f : \omega \rightarrow a$ such that $f(0) = b_0$ and $(\forall n{\in}\omega)\big[\phi(f(n)) \wedge \psi(f(n), f(n+1))\big]$. The latter principle is weaker than our Δ_0-RDC as all quantifiers have to be restricted to a given set a. However, the realizability interpretation of constructive set theory in PA$^r_\Omega$ + Σ^Ω-IND employed in the proof of ([27]: Theorem 3.15 (i)) also validates the stronger Δ_0-RDC of the present paper (the system PA$^r_\Omega$ stems from [17]).

Theorem 3.15 (i) of [27] and Lemma 3.8 also imply that CZF$^-$ + Δ_0-RDC is not weaker than CZF$^-$ + Σ_1-IND$_\omega$. Thus proof-theoretic equivalence of all systems in (i) ensues.

(iii) is a consequence of remark 3.3. At present the exact proof-theoretic strength of CZFA is not known, however, it can be shown that the proof-theoretic ordinal of CZFA is not bigger than $\varphi 20$. The latter bound can be obtained by inspecting the interpretation of CZFA in $\mathrm{PA}^r_\Omega + \Sigma^\Omega$-IND employed in the proof of [27]: Theorem 3.15. A careful inspection reveals that a subtheory T of $\mathrm{PA}^r_\Omega + \Sigma^\Omega$-IND suffices. To be more precise, T can be taken to be the theory

$$\mathrm{PA}^r_\Omega + \forall\alpha\,\exists\lambda\,[\alpha < \lambda \ \wedge\ \lambda \text{ is a limit ordinal}].$$

Using cut elimination techniques and asymmetric interpretation, T can be partially interpreted in $\mathrm{RA}_{<\omega^2}$. The latter theory is known to have proof-theoretic ordinal $\varphi 20$.

(iv) The above ordinals are less than Γ_0. □

Remark 4.2. Constructive set theory with AFA has an interpretation in Martin-Löf type theory as has been shown by I. Lindström [21]. Martin-Löf type theory is considered to be the most acceptable foundational framework of ideas that makes precise the constructive approach to mathematics. The interpretation of CZFA in Martin-Löf type theory demonstrates that there is a constructive notion of set that lends constructive meaning to AFA. However, Martin-Löf type theory is not a predicative theory in the sense of Feferman and Schütte as it possesses a proof-theoretic ordinal bigger than Γ_0. The work in [27] shows that CZFA and its variants can also be reduced to theories which are predicative in the stricter sense of autonomous progressions.

5. On Using the Anti-Foundation Axiom

In this section I rummage through several applications of AFA made in [3] and [5]. In order to corroborate my claim that most applications of AFA require only constructive means, various sections of [3] and [5] are recast on the basis of the theory CZFA rather than ZFA.

5.1. The Labelled Anti-Foundation Axiom

In applications it is often useful to have a more general form of AFA at one's disposal.

Definition 5.1. A *labelled graph* is a graph together with a labelling function ℓ which assigns a set $\ell(a)$ of *labels* to each node a.

A *labelled decoration* of a labelled graph is a function d such that

$$d(a) = \{d(b) : a \to b\} \cup \ell(a).$$

An unlabelled graph $(G, \twoheadrightarrow)$ may be identified with the special labelled graph where the labelling function $\ell : G \to V$ always assigns the empty set, i.e., $\ell(x) = \emptyset$ for all $x \in G$.

Theorem 5.2 ((CZFA), cf. [3]: Theorem 1.9). *Each labelled graph has a unique labelled decoration.*

Proof. Let $\mathbb{G} = (G, \twoheadrightarrow, \ell)$ be a labelled graph. Let $\mathbb{G}' = (G', \rightarrow)$ be the graph having as nodes all the ordered pairs $\langle i, a\rangle$ such that either $i = 1$ and $a \in G$ or $i = 2$ and $a \in \mathrm{TC}(G)$ and having as edges:

- $\langle 1, a\rangle \rightarrow \langle 1, b\rangle$ whenever $a \twoheadrightarrow b$,
- $\langle 1, a\rangle \rightarrow \langle 2, b\rangle$ whenever $a \in G$ and $b \in \ell(a)$,
- $\langle 2, a\rangle \rightarrow \langle 2, b\rangle$ whenever $b \in a \in \mathrm{TC}(G)$.

By AFA, $\mathbb{G}'$ has a unique decoration π. So for each $a \in G$

$$\pi(\langle 1, a\rangle) = \{\pi(\langle 1, b\rangle) : a \twoheadrightarrow b\} \cup \{\pi(\langle 2, b\rangle) : b \in \ell(a)\}$$

and for each $a \in \mathrm{TC}(G)$,

$$\pi(\langle 2, a\rangle) = \{\pi(\langle 2, b\rangle) : b \in a\}.$$

Note that the set $\mathrm{TC}(G)$ is naturally equipped with a graph structure by letting its edges $x \multimap y$ be defined by $y \in x$. The unique decoration for $(\mathrm{TC}(G), \multimap)$ is obviously the identity function on $\mathrm{TC}(G)$. As $x \mapsto \pi(\langle 2, x\rangle)$ is also a decoration of $(\mathrm{TC}(G), \multimap)$ we can conclude that $\pi(\langle 2, x\rangle) = x$ holds for all $x \in \mathrm{TC}(G)$. Hence if we let $\tau(a) = \pi(\langle 1, a\rangle)$ for $a \in G$ then, for $a \in G$,

$$\tau(a) = \{\tau(b) : a \twoheadrightarrow b\} \cup \ell(a),$$

so that τ is a labelled decoration of the labelled graph $\mathbb{G}$.

For the uniqueness of τ suppose that τ' is a labelled decoration of $\mathbb{G}$. Then π' is a decoration of the graph $\mathbb{G}'$, where

$$\pi'(\langle 1, a\rangle) = \tau'(a) \text{ for } a \in G,$$
$$\pi'(\langle 2, a\rangle) = a \text{ for } a \in \mathrm{TC}(G).$$

It follows from AFA that $\pi' = \pi$ so that for $a \in G$

$$\tau'(a) = \pi'(\langle 1, a\rangle) = \pi(\langle 1, a\rangle) = \tau(a),$$

and hence $\tau' = \tau$. □

Definition 5.3. A relation R is a *bisimulation* between two labelled graphs $\mathbb{G} = (G, \twoheadrightarrow, \ell_0)$ and $\mathbb{H} = (H, \rightarrowtail, \ell_1)$ if $R \subseteq G \times H$ and the following conditions are satisfied (where aRb stands for $\langle a, b\rangle \in R$):

1. For every $a \in G$ there is a $b \in H$ such that aRb.
2. For every $b \in H$ there is a $a \in G$ such that aRb.

3. Suppose that aRb. Then for every $x \in G$ such that $a \twoheadrightarrow x$ there is a $y \in H$ such that $b \rightarrowtail y$ and xRy.

4. Suppose that aRb. Then for every $y \in H$ such that $b \rightarrowtail y$ there is an $x \in G$ such that $a \twoheadrightarrow x$ and xRy.

5. If aRb then $\ell_0(a) = \ell_1(b)$.

Two labelled graphs are *bisimular* if there exists a bisimulation between them.

Theorem 5.4 (CZFA). *Let $\mathbb{G} = (G, \twoheadrightarrow, \ell_0)$ and $\mathbb{H} = (H, \rightarrowtail, \ell_1)$ be labelled graphs with labelled decorations d_0 and d_1, respectively.*
If $\mathbb{G}$ and $\mathbb{H}$ are bisimular then $d_0[G] = d_1[H]$.

Proof. Define a labelled graph $\mathbb{K} = (K, \to, \ell)$ by letting K be the set $\{\langle a, b\rangle : aRb\}$. For $\langle a, b\rangle, \langle a', b'\rangle \in K$ let $\langle a, b\rangle \to \langle a', b'\rangle$ iff $a \twoheadrightarrow a'$ or $b \rightarrowtail b'$, and put $\ell(\langle a, b\rangle) = \ell_0(a) = \ell_1(b)$. $\mathbb{K}$ has a unique labelled decoration d. Using a bisimulation R, one easily verifies that $d_0^*(\langle a, b\rangle) := d_0(a)$ and $d_1^*(\langle a, b\rangle) := d_1(b)$ are labelled decorations of $\mathbb{K}$ as well. Hence $d = d_0^* = d_1^*$, and thus $d_0[G] = d[K] = d_1[H]$. □

Corollary 5.5 (CZFA). *Two graphs are bisimular if and only if their decorations have the same image.*

Proof. One direction follows from the previous theorem. Now suppose we have graphs $\mathbb{G} = (G, \twoheadrightarrow)$ and $\mathbb{H} = (H, \rightarrowtail)$ with decorations d_0 and d_1, respectively, such that $d_0[G] = d_1[H]$. Then define $R \subseteq G \times H$ by aRb iff $d_0(a) = d_1(b)$. One readily verifies that R is a bisimulation. □

Here is another useful fact:

Lemma 5.6 (CZFA). *If A is transitive set and $d : A \to V$ is a function such that $d(a) = \{d(x) : x \in a\}$ for all $a \in A$, then $d(a) = a$ for all $a \in A$.*

Proof. A can be considered the set of nodes of the graph $\mathbb{G}_A = (A, \rightarrowtail)$ where $a \rightarrowtail b$ iff $b \in a$ and $a, b \in A$. Since A is transitive, d is a decoration of $\mathbb{G}$. But so is the function $a \mapsto a$. Thus we get $d(a) = a$. □

5.2. Systems

In applications it is often useful to avail oneself of graphs that are classes rather than sets. By a *map* $\wp$ with domain M we mean a definable class function with domain M, and we will write $\wp : M \to V$.

Definition 5.7. A *labelled system* is a class M of nodes together with a labelling map $\wp : M \to V$ and a class E of edges consisting of ordered pairs of nodes. Furthermore, a system is required to satisfy that for each node $a \in M$, $\{b \in M : a \rightarrowtail b\}$ is a set, where $a \rightarrowtail b$ stands for $\langle a, b\rangle \in E$.

The labelled system is said to be Δ_0 if the relation between sets x and y defined by "$y = \{b \in M : a \rightarrowtail b \text{ for some } a \in x\}$" is Δ_0 definable.

We will abbreviate the labelled system by $\mathbb{M} = (M, \rightarrowtail, \wp)$.

Theorem 5.8 ((CZFA + IND_ω), cf. [3]: Theorem 1.10). *For every labelled system $\mathbb{M} = (M, \rightarrowtail, \wp)$ there exists a unique map $d : M \to V$ such that, for all $a \in M$:*

$$d(a) = \{d(b) : a \to b\} \cup \ell(a). \tag{2}$$

Proof. To each $a \in M$ we may associate a labelled graph $\mathbb{M}_a = (M_a, {}_a\!\rightarrowtail, \wp_a)$ with $M_a = \bigcup_{n\in\omega} X_n$, where $X_0 = \{a\}$ and $X_{n+1} = \{b : a \rightarrowtail b \text{ for some } a \in X_n\}$. The existence of the function $n \mapsto X_n$ is shown via recursion on ω, utilizing IND_ω in combination with Strong Collection. The latter is needed to show that for every set Y, $\{b : a \rightarrowtail b \text{ for some } a \in Y\}$ is a set as well. And consequently to that M_a is a set. ${}_a\!\rightarrowtail$ is the restriction of $\rightarrowtail$ to nodes from M_a. That $E_a = \{\langle x, y\rangle \in M_a \times M_a : x \rightarrowtail y\}$ is a set requires Strong Collection, too. Further, let $\wp_a$ be the restriction of $\wp$ to M_a. Hence $\mathbb{M}_a$ is a set and we may apply Theorem 5.2 to conclude that $\mathbb{M}_a$ has a unique labelled decoration d_a. $d : M \to V$ is now obtained by patching together the function d_a with $a \in M$, that is $d = \bigcup_{a\in V} d_a$. One easily shows that two function d_a and d_b agree on $M_a \cap M_b$. For the uniqueness of d, notice that every other definable map d' satisfying (2) yields a function when restricted to M_a (Strong Collection) and thereby yields also a labelled decoration of $\mathbb{M}_a$; thus $d'(x) = \wp_a(x) = d(x)$ for all $x \in M_a$. And consequently to that, $d'(x) = d(x)$ for all $x \in M$. □

Corollary 5.9 (CZFA + Σ-IND_ω). *For every labelled system $\mathbb{M} = (M, \rightarrowtail, \wp)$ that is Δ_0 there exists a unique map $d : M \to V$ such that, for all $a \in M$:*

$$d(a) = \{d(b) : a \to b\} \cup \ell(a). \tag{3}$$

Proof. This follows by scrutinizing the proof of Theorem 5.8 and realizing that for a Δ_0 system one only needs Σ-IND_ω. □

Corollary 5.10 (CZFA). *Let $\mathbb{M} = (M, \rightarrowtail, \wp)$ be a labelled Δ_0 system such that for each $a \in M$ there is a function $n \mapsto X_n$ with domain ω such that $X_0 = \{a\}$ and $X_{n+1} = \{b : a \rightarrowtail b \text{ for some } a \in X_n\}$. Then there exists a unique map $d : M \to V$ such that, for all $a \in M$:*

$$d(a) = \{d(b) : a \to b\} \cup \ell(a). \tag{4}$$

Proof. In the proof of Theorem 5.8 we employed IND_ω only once to ensure that $M_a = \bigcup_{n\in\omega} X_n$ is a set. This we get now for free from the assumptions. □

Theorem 5.11 ((CZFA + IND_ω), cf. [3]: Theorem 1.11). *Let $\mathbb{M} = (M, \twoheadrightarrow, \wp)$ be a labelled system whose sets of labels are subsets of the class Y.*

1. *If π is a map with domain Y then there is a unique function $\hat{\pi}$ with domain M such that for each $a \in M$*

$$\hat{\pi}(a) = \{\hat{\pi}(b) : a \twoheadrightarrow b\} \cup \{\pi(x) : x \in \wp(a)\}.$$

2. *Given a map $\hbar : Y \to M$, there is a unique map π with domain Y such that for all $y \in Y$,*

$$\pi(y) = \hat{\pi}(\hbar(y)).$$

Proof. For (1) let $\mathbb{M}_\pi = (M, \twoheadrightarrow, \wp_\pi)$ be obtained from $\mathbb{M}$ and $\pi : Y \to V$ by redefining the sets of labels so that for each node a

$$\wp_\pi(a) = \{\pi(x) : x \in \wp(a)\}.$$

Then the required unique map $\hat{\pi}$ is the unique labelled decoration of $\mathbb{M}_\pi$ provided by Theorem 5.8

For (2) let $\mathbb{M}^* = (M, \rightarrowtail)$ be the graph having the same nodes as $\mathbb{M}$, and all edges of $\mathbb{M}$ together with the edges $a \rightarrowtail \hbar(y)$ whenever $a \in M$ and $y \in \wp(a)$. By Theorem 5.8, $\mathbb{M}^*$ has a unique decoration map ρ. So for each $a \in M$

$$\rho(a) = \{\rho(b) : a \twoheadrightarrow b\} \cup \{\rho(\hbar(y)) : y \in \wp(a)\}.$$

Letting $\pi(y) := \rho(\hbar(y))$ for $y \in Y$, ρ is also a labelled decoration for the labelled system $\mathbb{M}_\pi$ so that $\rho = \hat{\pi}$ by (1), and hence $\pi(x) = \hat{\pi}(\hbar(x))$ for $x \in Y$. For the uniqueness of π let $\mu : M \to V$ satisfy $\mu(x) = \hat{\mu}(\hbar(x))$ for $x \in Y$. Then $\hat{\mu}$ is a decoration of $\mathbb{M}^*$ as well, so that $\hat{\mu} = \rho$. As a result $\mu(x) = \hat{\mu}(\hbar(x)) = \rho(\hbar(x)) = \pi(x)$ for $x \in Y$. Thus $\mu(x) = \pi(x)$ for all $x \in Y$. □

Corollary 5.12 (CZFA + Σ-IND_ω). *Let $\mathbb{M} = (M, \twoheadrightarrow, \wp)$ be a labelled system that is Δ_0 and whose sets of labels are subsets of the class Y.*

1. *If π is a map with domain Y then there is a unique function $\hat{\pi}$ with domain M such that for each $a \in M$*

$$\hat{\pi}(a) = \{\hat{\pi}(b) : a \twoheadrightarrow b\} \cup \{\pi(x) : x \in \wp(a)\}.$$

2. *Given a map $\hbar : Y \to M$ there is a unique map π with domain Y such that for all $x \in Y$,*

$$\pi(x) = \hat{\pi}(\hbar(x)).$$

Proof. The proof is the same as for Theorem 5.11, except that one utilizes Corollary 5.9 in place of Theorem 5.8. □

Corollary 5.13 (CZFA). *Let $\mathbb{M} = (M, \twoheadrightarrow, \wp)$ be a labelled system that is Δ_0 and whose sets of labels are subsets of the class Y. Moreover suppose that for each $a \in M$ there is a function $n \mapsto X_n$ with domain ω such that $X_0 = \{a\}$ and $X_{n+1} = \{b : a \rightarrowtail b \text{ for some } a \in X_n\}$.*

1. *If π is a map with domain Y then there is a unique function $\hat{\pi}$ with domain M such that for each $a \in M$*

$$\hat{\pi}(a) = \{\hat{\pi}(b) : a \twoheadrightarrow b\} \cup \{\pi(x) : x \in \wp(a)\}.$$

2. *Given a map $\hbar : Y \to M$ there is a unique map π with domain Y such that for all $x \in Y$,*

$$\pi(x) = \hat{\pi}(\hbar(x)).$$

Proof. The proof is the same as for Theorem 5.11, except that one utilizes Corollary 5.10 in place of Theorem 5.8. □

5.3. A Solution Lemma Version of AFA

AFA can be couched in more traditional mathematical terms. The labelled Anti-Foundation Axiom provides a nice tool for showing that systems of equations of a certain type have always unique solutions. In the terminology of [5] this is called the *solution lemma*. In [5], the Anti-Foundation Axiom is even expressed in terms of unique solutions to so-called *flat systems of equations*.

Definition 5.14. For a set Y let $\mathcal{P}(Y)$ be the class of subsets of Y. A triple $\mathcal{E} = (X, A, e)$ is said to be a *general flat system of equations* if X and A are any two sets, and $e : X \to \mathcal{P}(X \cup A)$, where the latter conveys that e is a function with domain X which maps into the class of all subsets of $X \cup A$. X will be called the set of *indeterminates* of $\mathcal{E}$, and A is called the set of *atoms* of $\mathcal{E}$. Let $e_v = e(v)$. For each $v \in X$, the set $b_v := e_v \cap X$ is called the set of indeterminates on which v immediately depends. Similarly, the set $c_v := e_v \cap A$ is called the set of atoms on which v immediately depends.

A *solution* to $\mathcal{E}$ is a function s with domain X satisfying

$$s_x = \{s_y : y \in b_x\} \cup c_x,$$

for each $x \in X$, where $s_x := s(x)$.

Theorem 5.15 (CZFA). *Every generalized flat system $\mathcal{E} = (X, A, e)$ has a unique solution.*

Proof. Define a labelled graph $\mathbb{H}$ by letting X be its set of nodes and its edges be of the form $x \twoheadrightarrow y$, where $y \in b_x$ for $x, y \in X$. Moreover, let $\ell(x) = c_x$ be the pertinent

labelling function. By Theorem 5.2, $\mathbb{H}$ has a unique labelled decoration d. Then

$$d(x) = \{d(y) : y \in b_x\} \cup \ell(x) = \{d(y) : y \in b_x\} \cup c_x,$$

and thus d is a solution to $\mathcal{E}$. One easily verifies that every solution s to $\mathcal{E}$ gives rise to a decoration of $\mathbb{H}$. Thus there exists exactly one solution to $\mathcal{E}$. □

Because of the flatness condition, i.e., $e : X \to \mathcal{P}(X \cup A)$, the above form of the Solution Lemma is often awkward to use. A much more general form of it is proved in [5]. The framework in [5], however, includes other objects than sets, namely a proper class of urelements, whose *raison d'etre* is to serve as an endless supply of indeterminates on which one can perform the operation of substitution. Given a set X of urelements one defines the class of X-sets which are those sets that use only urelements from X in their build-up. For a function $f : X \to V$ on these indeterminates one can then define a substitution operation sub_f on the X-sets. For an X-set a, $\mathrm{sub}_f(a)$ is obtained from a by substituting $f(x)$ for x everywhere in the build-up of a.

For want of urelements, the approach of [5] is not directly applicable in our set theories, though it is possible to model an extended universe of sets with a proper class of urelements within CZFA. This will require a class defined as the greatest fixed point of an operator, a topic I shall intersperse now.

5.4. Greatest Fixed Points of Operators

The theory of greatest greatest fixed points was initiated by Aczel in [3].

Definition 5.16. Let Φ be a class operator, i.e. $\Phi(X)$ is a class for each class X. Φ is *set continuous* if for each class X

$$\Phi(X) = \bigcup\{\Phi(x) : x \text{ is a set with } x \subseteq X\}. \tag{5}$$

Note that a set continuous operator is monotone, i.e., if $X \subseteq Y$ then $\Phi(X) \subseteq \Phi(Y)$.

In what follows, I shall convey that x is a set by $x \in V$. If Φ is a set continuous operator let

$$J_\Phi = \bigcup\{x \in V : x \subseteq \Phi(x)\}.$$

A set continuous operator Φ is Δ_0 if the relation "$y \in \Phi(x)$" between sets x and y is Δ_0 definable. Notice that J_Φ is a Σ_1 class if Φ is a Δ_0 operator.

Theorem 5.17 ((CZF$^-$ + RDC), cf. [3]: Theorem 6.5). *If Φ is a set continuous operator and $J = J_\Phi$ then*

1. $J \subseteq \Phi(J)$,
2. *If $X \subseteq \Phi(X)$ then $X \subseteq J$,*

3. *J is the largest fixed point of Φ.*

Proof. (1): Let $a \in J$. Then $a \in x$ for some set x such that $x \subseteq \Phi(x)$. It follows that $a \in \Phi(J)$ as $x \subseteq J$ and Φ is monotone.

(2): Let $X \subseteq \Phi(X)$ and let $a \in X$. We like to show that $a \in J$. We first show that for each set $x \subseteq X$ there is a set $c_x \subseteq X$ such that $x \subseteq \Phi(c_x)$. So let $x \subseteq X$. Then $x \subseteq \Phi(X)$ yielding

$$\forall y \in x\ \exists u\ [y \in \Phi(u)\ \wedge\ u \subseteq X].$$

By Strong Collection there is a set A such that

$$\forall y \in x\ \exists u \in A\ [y \in \Phi(u)\ \wedge\ u \subseteq X]\ \wedge\ \forall u \in A\ \exists y \in x\ [y \in \Phi(u)\ \wedge\ u \subseteq X].$$

Letting $c_x = \bigcup A$, we get $c_x \subseteq X\ \wedge\ x \subseteq \Phi(c_x)$ as required.

Next we use RDC to find an infinite sequence $x_0, x_1, \ldots$ of subsets of X such that $x_0 = \{a\}$ and $x_n \subseteq \Phi(x_{n+1})$. Let $x^* = \bigcup_n x_n$. Then x^* is a set and if $y \in x^*$ then $y \in x_n$ for some n so that $y \in x_n \subseteq \Phi(x_{n+1}) \subseteq \Phi(x^*)$. Hence $x^* \subseteq \Phi(x^*)$. As $a \in x_0 \subseteq x^*$ it follows that $a \in J$.

(3): By (1) and the monotonicity of Φ

$$\Phi(J) \subseteq \Phi(\Phi(J)).$$

Hence by (2) $\Phi(J) \subseteq J$. This and (1) imply that J is a fixed point of Φ. By (2) it must be the greatest fixed point of Φ. □

If it exists and is a set, the largest fixed point of an operator Φ will be called the set *coinductively defined* by Φ.

Theorem 5.18 ($\mathrm{CZF}^- + \Delta_0$-RDC). *If Φ is a set continuous Δ_0 operator and $J = J_\Phi$ then*

1. *$J \subseteq \Phi(J)$,*
2. *If X is a Σ_1 class and $X \subseteq \Phi(X)$ then $X \subseteq J$,*
3. *J is the largest Σ_1 fixed point of Φ.*

Proof. This is the same proof as for Theorem 5.17, noticing that Δ_0-RDC suffices here. □

In applications, set continuous operators Φ often satisfy an additional property. Φ will be called *fathomable* if there is a partial class function q such that whenever $a \in \Phi(x)$ for some set x then $q(a) \subseteq x$ and $a \in \Phi(q(a))$. For example, deterministic inductive definitions are given by fathomable operators.

If the graph of q is also Δ_0 definable we will say that Φ is a fathomable set continuous Δ_0 operator.

For fathomable operators one can dispense with RDC and Δ_0-RDC in Theorems 5.17 and 5.18 in favour of IND_ω and Σ-IND_ω, respectively.

Corollary 5.19 ($\mathrm{CZF}^- + \mathrm{IND}_\omega$). *If Φ is a set continuous fathomable operator and $J = J_\Phi$ then*

1. $J \subseteq \Phi(J)$,
2. *If $X \subseteq \Phi(X)$ then $X \subseteq J$,*
3. *J is the largest fixed point of Φ.*

Proof. In the proof of Theorem 5.17, RDC was used for (2) to show that for every class X with $X \subseteq \Phi(X)$ it holds $X \subseteq J$. Now, if $a \in X$, then $a \in \Phi(u)$ for some set $u \subseteq X$, as Φ is set continuous, and thus $a \in \Phi(q(a))$ and $q(a) \subseteq X$. Using IND_ω and Strong Collection one defines a sequence $x_0, x_1, \ldots$ by $x_0 = \{a\}$ and $x_{n+1} = \bigcup\{q(v) : v \in x_n\}$. We use induction on ω to show $x_n \subseteq X$. Obviously $x_0 \subseteq X$. Suppose $x_n \subseteq X$. Then $x_n \subseteq \Phi(X)$. Thus for every $v \in x_n$, $q(v) \subseteq X$, and hence $x_{n+1} \subseteq X$. Let $x^* = \bigcup_n x_n$. Then $x^* \subseteq X$. Suppose $u \in x^*$. Then $u \in x_n$ for some n, and hence as $u \in \Phi(X)$, $u \in \Phi(q(u))$. Thus $q(u) \subseteq x_{n+1} \subseteq x^*$, and so $u \in \Phi(x^*)$. As a result, $a \in x^* \subseteq \Phi(x^*)$, and hence $a \in J$. □

Corollary 5.20 ($\mathrm{CZF}^- + \Sigma\text{-}\mathrm{IND}_\omega$). *If Φ is a set continuous fathomable Δ_0 operator and $J = J_\Phi$ then*

1. $J \subseteq \Phi(J)$,
2. *If X is Σ_1 and $X \subseteq \Phi(X)$ then $X \subseteq J$,*
3. *J is the largest Σ_1 fixed point of Φ.*

Proof. If the graph of q is Δ_0 definable, $\Sigma\text{-}\mathrm{IND}_\omega$ is sufficient to define the sequence $x_0, x_1, \ldots$. □

For special operators it is also possible to forgo $\Sigma\text{-}\mathrm{IND}_\omega$ in favour of TRANS.

Corollary 5.21 ($\mathrm{CZF}^- + \mathrm{TRANS}$). *Let Φ be a set continuous fathomable Δ_0 operator such that q is a total map and $q(a) \subseteq \mathrm{TC}(\{a\})$ for all sets a. Let $J = J_\Phi$. Then*

1. $J \subseteq \Phi(J)$,
2. *If X is Δ_0 and $X \subseteq \Phi(X)$ then $X \subseteq J$,*
3. *J is the largest Δ_0 fixed point of Φ.*

Proof. (1) is proved as in Theorem 5.17. For (2), suppose that X is a class with $X \subseteq \Phi(X)$. Let $a \in X$. Define a sequence of sets $x_0, x_1, \ldots$ by $x_0 = \{a\}$ and $x_{n+1} = \bigcup\{q(v) : v \in x_n\}$ as in Corollary 5.19. But without $\Sigma\text{-}\mathrm{IND}_\omega$, how can we ensure that the function $n \mapsto x_n$ exists? This can be seen as follows. Define

$$D_n = \{f \in {}^{n+1}\mathrm{TC}(\{a\}) : f(0) = a \wedge \forall i \in n\, [f(i+1) \in q(f(i))]\},$$
$$E_n = \{f(n) : f \in D_n\}.$$

The function $n \mapsto E_n$ exists by Strong Collection. Moreover, $E_0 = \{a\}$ and $E_{n+1} = \bigcup\{q(v) : v \in E_n\}$ as can be easily shown by induction on n; thus $x_n = E_n$. The remainder of the proof is as in Corollary 5.19.

For (3), note that $J = \{a : a \in \Phi(q(a))\}$ and thus J is Δ_0. □

Remark 5.22. It is an open problem whether the above applications of the dependent choices axiom are necessary for the general theory of greatest fixed points.

5.5. Generalized Systems of Equations in an Expanded Universe

Before we can state the notion of a general systems of equations we will have to emulate urelements and the sets built out of them in the set theory CZFA with pure sets. To this end we employ the machinery of greatest fixed points of the previous subsection. We will take the sets of the form $\langle 1, x\rangle$ to be the urelements and call them $*$-*urelements*. The class of $*$-urelements will be denoted by $\mathcal{U}$. Certain sets built from them will be called the $*$-*sets*. If $a = \langle 2, u\rangle$ let $a^* = u$. The elements of a^* will be called the $*$-elements of a. Let the $*$-sets be the largest class of sets of the form $a = \langle 2, u\rangle$ such that each $*$-element of a is either a $*$-urelement or else a $*$-set. To bring this under the heading of the previous subsection, define

$$\Phi^*(X) = \{\langle 2, u\rangle : \forall x{\in}u \left[(x{\in}X \wedge x \in \text{TWO}) \vee x \text{ is a } *\text{-urelement}\right]\},$$

where TWO is the class of all ordered pairs of the form $\langle 2, v\rangle$. Obviously, Φ^* is a set-continuous operator. That Φ^* is fathomable can be seen by letting

$$q(a) = \{v \in a^* : v \in \text{TWO}\}.$$

Notice also that Φ^* has a Δ_0 definition.

The $*$-sets are precisely the elements of J_{Φ^*}. Given a class of Z of $*$-urelements we will also define the class of Z-*sets* to be the largest class of $*$-sets such that every $*$-urelement in a Z-set is in Z. We will use the notation $V[Z]$ for the class of Z-sets.

Definition 5.23. A *general system of equations* is a pair $\mathcal{E} = (X, e)$ consisting of a set $X \subseteq \mathcal{U}$ (of indeterminates) and a function

$$e : X \to V[X].$$

The point of requiring e to take values in $V[X]$ is that thereby e is barred from taking $*$-urelements as values and that all the values of e are sets which use only $*$-urelements from X in their build-up. In consequence, one can define a substitution operation on the values of e.

Theorem 5.24 ((CZFA), Substitution Lemma). *Let Y be a Δ_0 class such that $Y \subseteq \mathcal{U}$. For each map $\rho : Y \to V$ there exists a unique operation* sub_ρ *that assigns to each*

$a \in V[Y]$ *a set* $\mathrm{sub}_\rho(a)$ *such that*

$$\mathrm{sub}_\rho(a) = \{\mathrm{sub}_\rho(x) : x \in a^* \cap V[Y]\} \cup \{\rho(x) : x \in a^* \cap Y\}. \tag{6}$$

Proof. The class $V[Y]$ forms the nodes of a labelled Δ_0 system $\mathbb{M}$ with edges $a \rightarrowtail b$ for $a, b \in V[Y]$ whenever $b \in a^*$, and labelling map $\wp(a) = a^* \cap Y$. By Corollary 5.13 there exists a unique map $\hat{\rho} : V[Y] \to V$ such that for each $a \in V[Y]$,

$$\hat{\rho}(a) = \{\hat{\rho}(x) : x \in a^* \cap V[Y]\} \cup \{\rho(x) : x \in a^* \cap Y\}. \tag{7}$$

Put $\mathrm{sub}_\rho(a) := \hat{\rho}(a)$. Then sub_ρ satisfies (6). Since the equation (7) uniquely determines $\hat{\rho}$ it follows that sub_ρ is uniquely determined as well. □

Definition 5.25. Let $\mathcal{E}$ be a general system of equations as in Definition 5.23. A *solution* to $\mathcal{E}$ is a function $s : X \to V$ satisfying, for all $x \in X$,

$$s(x) = \mathrm{sub}_s(e_x), \tag{8}$$

where $e_x := e(x)$.

Theorem 5.26 ((CZFA), Solution Lemma). *Let $\mathcal{E}$ be a general system of equations as in Definition* 5.23. *Then $\mathcal{E}$ has a unique solution.*

Proof. The class $V[X]$ provides the nodes for a labelled Δ_0 system $\mathbb{M}$ with edges $b \rightarrowtail c$ for $b, c \in V[X]$ whenever $c \in b^*$, and with a labelling map $\wp(b) = b^* \cap X$. Since $e : X \to V[X]$, we may employ Corollary 5.13 (with $Y = X$). Thus there is a unique function π and a unique function $\hat{\pi}$ such that

$$\pi(x) = \hat{\pi}(e_x) \tag{9}$$

for all $x \in X$, and

$$\hat{\pi}(a) = \{\hat{\pi}(b) :: b \in a^*\} \cup \{\pi(x) : x \in a^* \cap X\}. \tag{10}$$

In view of Theorem 5.24, we get $\hat{\pi} = \mathrm{sub}_\pi$ from (10). Thus letting $s := \pi$, (9) then yields the desired equation $s(x) = \mathrm{sub}_s(e_x)$ for all $x \in X$. Further, s is unique owing to the uniqueness of π in (9). □

Remark 5.27. The framework in which AFA is studied in [5] is a set theory with a proper class of urelements $\mathcal{U}$ that also features an *axiom of plenitude* which is the conjunction of the following sentences:

$$\forall a \forall b\, \mathrm{new}(a, b) \in \mathcal{U},$$
$$\forall a \forall a' \forall b \forall b' \left[\mathrm{new}(a, b) = \mathrm{new}(a', b') \to a = a' \wedge b = b'\right],$$
$$\forall a \forall b \left[b \subseteq \mathcal{U} \to \mathrm{new}(a, b) \notin b\right],$$

where new is a binary function symbol. It is natural to ask whether a version of CZFA with urelements and an axiom of plenitude would yield any extra strength. That such a theory is not stronger than CZFA can be easily seen by modelling the urelements and

sets of [5] inside CZFA by the $*$-urelements and the $*$-sets, respectively. To interpret the function symbol new define

$$\text{new}^*(a, b) := \langle 1, \langle a, \langle b, b^r \rangle \rangle \rangle,$$

where $b^r = \{r \in \text{TC}(b) : r \notin r\}$. Obviously, $\text{new}^*(a, b)$ is a $*$-urelement and new^* is injective. Moreover, $\text{new}^*(a, b) \in b$ would imply $\text{new}^*(a, b) \in \text{TC}(b)$ and thus $b^r \in \text{TC}(b)$. The latter yields the contradiction $b^r \notin b^r \wedge b^r \in b^r$. As a result, $\text{new}^*(a, b) \notin b$. Interpreting new by new^* thus validates the axiom of plenitude, too.

5.6. Streams, Coinduction, and Corecursion

In the following I shall demonstrate the important methods of coinduction and corecursion in a setting which is not too complicated but still demonstrates the general case in a nutshell. The presentation closely follows [5].

Let A be some set. By a *stream* over A we mean an ordered pair $s = \langle a, s' \rangle$ where $a \in A$ and s' is another stream. We think of a stream as being an element of A followed by another stream. Two important operations performed on streams s are taking the first element $1^{\text{st}}(s)$ which gives an element of A, and taking its second element $2^{\text{nd}}(s)$, which yields another stream. If we let A^∞ be the streams over A, then we would like to have

$$A^\infty = A \times A^\infty. \tag{11}$$

In set theory with the foundation axiom, equation (11) has only the solution $A = \emptyset$. With AFA, however, not only can one show that (11) has a solution different from $\emptyset$ but also that it has a largest solution, the latter being the largest fixed point of the operator $\Gamma_A(Z) = A \times Z$. This largest solution to Γ_A will be taken to be the set of streams over A and be denoted by A^∞, thus rendering A^∞ a *coinductive* set. Moreover, it will be shown that A^∞ possesses a "recursive" character despite the fact that there is no "base case". For instance, it will turn out that one can define a function

$$\text{zip} : A^\infty \times A^\infty \to A^\infty$$

such that for all $s, t \in A^\infty$

$$\text{zip}(s, t) = \langle 1st(s), \langle 1^{\text{st}}(t), \text{zip}(2^{\text{nd}}(s), 2^{\text{nd}}(t)) \rangle \rangle. \tag{12}$$

As its name suggests, zip acts like a zipper on two streams. The definition of zip in (12) is an example for definition by *corecursion* over a coinductive set.

Theorem 5.28 (CZFA). *For every set A there is a largest set Z such that $Z \subseteq A \times Z$. Moreover, Z satisfies $Z = A \times Z$, and if A is inhabited then so is Z.*

Proof. Let F be the set of functions from $\mathbb{N} := \omega$ to A. For each such f, we define another function $f^+ : \mathbb{N} \to \mathbb{N}$ by

$$f^+(n) = f(n + 1).$$

For each $f \in F$ let x_f be an indeterminate. We would like to solve the system of equations given by

$$x_f = \langle f(0), x_{f^+} \rangle.$$

Solving these equations is equivalent to solving the equations

$$\begin{aligned} x_f &= \{y_f, z_f\}; \\ y_f &= \{f(0)\} \\ z_f &= \{f(0), x_{f^+}\}, \end{aligned} \tag{13}$$

where y_f and z_f are further indeterminates. Note that $f(0)$ is an element of A. To be precise, let $x_f = \langle 0, f \rangle$, $y_f = \langle 1, f \rangle$, and $z_f = \langle 2, f \rangle$. Solving (13) amounts to the same as finding a labelled decoration for the labelled graph

$$\mathbb{S}_A = (S, \twoheadrightarrow, \ell)$$

whose set of nodes is

$$S = \{x_f : f \in F\} \cup \{y_f : f \in F\} \cup \{z_f : f \in F\}$$

and whose edges are given by $x_f \twoheadrightarrow y_f, x_f \twoheadrightarrow z_f, z_f \twoheadrightarrow x_{f^+}$. Moreover, the labelling function ℓ is defined by $\ell(x_f) = \emptyset$, $\ell(y_f) = \{f(0)\}$, $\ell(z_f) = \{f(0)\}$ for all $f \in F$. By the labelled Anti-Foundation Axiom, Theorem 5.2, $\mathbb{S}_A$ has a labelled decoration d and we thus get

$$d(x_f) = \langle f(0), d(x_{f^+}) \rangle. \tag{14}$$

Let $A^\infty = \{d(x_f) : f \in F\}$. By (14), we have $A^\infty \subseteq A \times A^\infty$. Thus A^∞ solves the equation $Z \subseteq A \times Z$.

To check that $A \times A^\infty \subseteq A^\infty$ holds also, let $a \in A$ and $t \in A^\infty$. By the definition of A^∞, $t = d(x_f)$ for some $f \in F$. Let $g : \mathbb{N} \to A$ be defined by $g(0) = a$ and $g(n+1) = f(n)$. Then $g^+ = f$, and thus $d(x_g) = \langle a, d(x_f) \rangle = \langle a, t \rangle$, so $\langle a, t \rangle \in A^\infty$.

If A contains an element a, then $f_a \in F$, where $f_a : \mathbb{N} \to A$ is defined by $f_a(n) = a$. Hence $d(x_{f_a}) \in A^\infty$, so A^∞ is inhabited, too.

Finally it remains to show that A^∞ is the largest set Z satisfying $Z \subseteq A \times Z$. So suppose that W is a set so that $W \subseteq A \times W$. Let $v \in W$. Define $f_v : \mathbb{N} \to A$ by

$$f_v(n) = 1^{\text{st}}(\text{sec}^n(v)),$$

where $\text{sec}^0(v) = v$ and $\text{sec}^{n+1}(v) = 2^{\text{nd}}(\text{sec}^n(v))$. Then $f_v \in F$, and so $d(x_{f_v}) \in A^\infty$. We claim that for all $v \in W$, $d(x_{f_v}) = v$. Notice first that for $w = 2^{\text{nd}}(v)$, we have $\text{sec}^n(w) = \text{sec}^{n+1}(v)$ for all $n \in \mathbb{N}$, and thus $f_w = (f_v)^+$. It follows that

$$\begin{aligned} d(x_{f_v}) &= \langle 1^{\text{st}}(v), d(x_{(f_v)^+}) \rangle \\ &= \langle 1^{\text{st}}(v), d(x_{f_w}) \rangle \\ &= \langle 1^{\text{st}}(v), d(x_{f_{2^{\text{nd}}(v)}}) \rangle. \end{aligned} \tag{15}$$

W gives rise to a labelled subgraph $\mathbb{T}$ of $\mathbb{S}$ whose set of nodes is

$$T := \{x_{f_v} : v \in W\} \cup \{y_{f_v} : v \in W\} \cup \{z_{f_v} : v \in W\},$$

and wherein the edges and the labelling function are obtained from $\mathbb{S}$ by restriction to nodes from T. The function d' with $d'(x_{f_v}) = v$, $d'(y_{f_v}) = \{1^{\text{st}}(v)\}$, and $d'(z_{f_v}) = \{1^{\text{st}}(v), 2^{\text{nd}}(v)\}$ is obviously a labelled decoration of $\mathbb{T}$. By (15), d restricted to T is a labelled decoration of $\mathbb{T}$ as well. So by Theorem 5.2, $v = d'(x_{f_v}) = d(x_{f_v})$ for all $v \in W$, and thus $W \subseteq A^\infty$. □

Remark 5.29. Rather than applying the labelled Anti-Foundation Axiom one can utilize the solution lemma for general systems of equations (Theorem 5.26) in the above proof of Theorem 5.28. To this end let $B = \mathrm{TC}(A)$, $x_f = \langle 1, \langle 0, f\rangle\rangle$ for $f \in F$ and $x_b = \langle 1, \langle 1, b\rangle\rangle$ for $b \in B$. Set $X := \{x_f : f \in F\} \cup \{x_b : b \in B\}$. Then $X \subseteq \mathcal{U}$ and $\{x_f : f \in F\} \cap \{x_b : b \in B\} = \emptyset$.

Next define the unordered $*$-pair by $\{c, d\}^* = \langle 2, \{c, d\}\rangle$ and the ordered $*$-pair by $\langle c, d\rangle^* = \{\{c\}^*, \{c, d\}^*\}^*$. Note that with $c, d \in V[X]$ one also has $\{c, d\}^*, \langle c, d\rangle^* \in V[X]$.

Let $\mathcal{E} = (X, e)$ be the general system of equations with $e(x_f) = \langle x_{f(0)}, x_{f^+}\rangle^*$ for $f \in F$ and $e(x_b) = \langle 2, \{x_u \;\; u \in b\}\rangle$ for $b \in B$. Then $e : X \to V[X]$. By Theorem 5.26 there is a unique function $s : X \to V$ such that

$$s(x_b) = \mathrm{sub}_s(e(x_b)) = \{s(x_u) : u \in b\} \quad \text{for } b \in B, \tag{16}$$

$$s(x_f) = \mathrm{sub}_s(e(x_f)) = \langle s(x_{f(0)}), s(x_{f^+})\rangle \quad \text{for } f \in F. \tag{17}$$

From (16) and Lemma 5.6 it follows $s(x_b) = b$ for all $b \in B$, and thus from (17) it ensues that $s(x_f) = \langle f(0), s(x_{f^+})\rangle$ for $f \in F$. From here on one can proceed further just as in the proof of Theorem 5.26.

As a corollary to Theorem 5.28 one gets the following *coinduction principle* for A^∞.

Corollary 5.30 (CZFA). *If a set Z satisfies $Z \subseteq A \times Z$, then $Z \subseteq A^\infty$.*

Proof. This follows from the fact that A^∞ is the largest such set. □

The pivotal property of inductively defined sets is that one can define functions on them by structural recursion. For coinductively defined sets one has a dual principle, *corecursion*, which allows one to define functions mapping into the coinductive set.

Theorem 5.31 ((CZFA), Corecursion Principle for Streams). *Let C be an arbitrary set. Given functions $g : C \to A$ and $h : C \to C$ there is a unique function $f : C \to A^\infty$ satisfying*

$$f(c) = \langle g(c), f(h(c))\rangle \tag{18}$$

for all $c \in C$.

Proof. For each $c \in C$ let x_c, y_c, z_c be different indeterminates. To be precise, let $x_c = \langle 0, c\rangle$, $y_c = \langle 1, c\rangle$, and $z_c = \langle 2, c\rangle$ for $c \in C$. This time we would like to solve the system of equations given by

$$x_c \;=\; \langle g(c), x_{h(c)}\rangle.$$

Solving these equations is equivalent to solving the equations

$$\begin{aligned} x_c &= \{y_c, z_c\};\\ y_c &= \{g(c)\};\\ z_c &= \{g(c), x_{h(c)}\}. \end{aligned} \tag{19}$$

Solving (19) amounts to the same as finding a labelled decoration for the labelled graph

$$\mathbb{S}_C = (S_C, \twoheadrightarrow, \ell_C)$$

whose set of nodes is

$$S_C = \{x_c \,:\, c \in C\} \cup \{y_c \,:\, c \in C\} \cup \{z_c \,:\, c \in C\}$$

and whose edges are given by $x_c \twoheadrightarrow y_c, x_c \twoheadrightarrow z_c, z_c \twoheadrightarrow x_{h(c)}$. Moreover, the labelling function ℓ_C is defined by $\ell_C(x_b) = \emptyset$, $\ell_C(y_b) = \{g(b)\}$, $\ell_C(z_b) = \{g(b)\}$ for all $b \in C$. By the labelled Anti-Foundation Axiom, Theorem 5.2, $\mathbb{S}_C$ has a labelled decoration $\jmath$ and we thus get

$$\jmath(x_c) \;=\; \langle g(c), \jmath(x_{h(c)})\rangle. \tag{20}$$

Letting the function f with domain C be defined by $f(c) := \jmath(x_c)$, we get from (20) that

$$f(c) \;=\; \langle g(c), f(h(c))\rangle \tag{21}$$

holds for all $c \in C$. As $\mathrm{ran}(f) \subseteq A \times \mathrm{ran}(f)$, Corollary 5.30 yields $\mathrm{ran}(f) \subseteq A^\infty$, thus $f : C \to A^\infty$.

It remains to show that f is uniquely determined by (21). So suppose $f' : C \to A^\infty$ is another function satisfying $f'(c) = \langle g(c), f'(h(c))\rangle$ for all $c \in C$. Then the function $\jmath'$ with $\jmath'(x_c) = f'(c)$, $\jmath'(y_c) = \{g(c)\}$, and $\jmath'(z_c) = \{g(c), f'(h(c))\}$ would give another labelled decoration of $\mathbb{S}_C$, hence $f(c) = \jmath(x_c) = \jmath'(x_c) = f'(x_c)$, yielding $f = f'$. □

Example 1. Let $k : A \to A$ be arbitrary. Then k gives rise to a unique function $\mathrm{map}_k : A^\infty \to A^\infty$ satisfying

$$\mathrm{map}_k(s) \;=\; \langle k(1^{\mathrm{st}}(s)), \mathrm{map}_k(2^{\mathrm{nd}}(s))\rangle. \tag{22}$$

For example, if $A = \mathbb{N}$, $k(n) = 2n$, and $s = \langle 3, \langle 6, \langle 9, \ldots\rangle\rangle\rangle$, then $\mathrm{map}_k(s) = \langle 6, \langle 12, \langle 18, \ldots\rangle\rangle\rangle$. To see that map_k exists, let $C = A^\infty$ in Theorem 5.31, $g : A^\infty \to A$ be defined by $g(s) = k(1^{\mathrm{st}}(s))$, and $h : A^\infty \to A^\infty$ be the function $h(s) = 2^{\mathrm{nd}}(s)$. Then map_k is the unique function f provided by Theorem 5.31.

Example 2. Let $\nu : A \to A$. We want to define a function

$$\mathrm{iter}_\nu \ : \ A \to A^\infty$$

which "iterates" ν such that $\mathrm{iter}_\nu(a) = \langle a, \mathrm{iter}_\nu(\nu(a))\rangle$ for all $a \in A$. If, for example $A = \mathbb{N}$ and $\nu(n) = 2n$, then $\mathrm{iter}_\nu(7) = \langle 7, \langle 14, \langle 28, \ldots\rangle\rangle\rangle$. To arrive at iter_ν we employ Theorem 5.31 with $C = A^\infty$, $g : C \to A$, and $h : C \to C$, where $g(s) = \nu(1^{\mathrm{st}}(s))$ and $h = \mathrm{map}_\nu$, respectively.

Outlook. It would be desirable to develop the theory of corecursion of [5] (in particular Theorem 17.5) and the final coalgebra theorem of [3] in full generality within CZFA and extensions. It appears that the first challenge here is to formalize parts of category theory in constructive set theory. Due to page restrictions this cannot not be done in the present paper.

Acknowledgement. I am grateful to the Mittag-Leffler Institute (Djursholm, Sweden) for giving me the opportunity in 2001 to be working on the talk on which this paper is based. The research reported in this paper was also partly supported by United Kingdom Engineering and Physical Sciences Research Council Grant GR/R 15856/01.

References

[1] Aczel, Peter: 1978. The type theoretic interpretation of constructive set theory. In: A. MacIntyre *et al*. (eds.), *Logic Colloquium '77*, Amsterdam: North-Holland, 55–66.

[2] Aczel, Peter: 1982. The type theoretic interpretation of constructive set theory: choice principles. In: A. S. Troelstra *et al*. (eds.), *The L. E. J. Brouwer Centenary Symposium*, Amsterdam: North–Holland, 1–40.

[3] Aczel, Peter: 1988. *Non-Well-Founded Sets*. CSLI Lecture Notes 14. Stanford: CSLI Publications.

[4] Barwise, Jon: 1975. *Admissible Sets and Structures. An Approach to Definability Theory*. Berlin: Springer.

[5] Barwise, Jon and Lawrence Moss: 1996. *Vicious Circles*. CSLI Lecture Notes 60. Stanford: CSLI Publications.

[6] Beeson, Michael: 1985. *Foundations of Constructive Mathematics*. Berlin: Springer.

[7] Cantini, Andrea: 1986. On the relation between choice and comprehension principles in second order arithmetic. *Journal of Symbolic Logic* vol. 51: 360–373.

[8] Crosilla, Laura: 1998. *Realizability Interpretations for Subsystems of* CZF *and Proof Theoretic Strength*. Technical Report. School of Mathematics, University of Leeds.

[9] Crosilla, Laura: 2000. *Realizability Models for Constructive Set Theories with Restricted Induction*. PhD Thesis. School of Mathematics, University of Leeds.

[10] Crosilla, Laura and Michael Rathjen: 2002. Inaccessible set axioms may have little consistency strength. *Annals of Pure and Applied Logic* 115: 33–70.

[11] Feferman, Solomon: 1964. Systems of predicative analysis I. *Journal of Symbolic Logic* 29: 1–30.

[12] Feferman, Solomon: 1968. Systems of predicative analysis II. Representations of ordinals. *Journal of Symbolic Logic* 33: 193–220.

[13] Feferman, Solomon: 1975. A language and axioms for explicit mathematics. In: J. N. Crossley (ed.), *Algebra and Logic*, Lecture Notes in Math. 450, Berlin: Springer: 87–139.

[14] Feferman, Solomon: 1977. Theories of finite type related to mathematical practice. In: J. Barwise (ed.): *Handbook of Mathematical Logic*, Amsterdam: North Holland, 913–971.

[15] Hallnäs, Lars: 1986. *Non Wellfounded Sets: Limits of Wellfounded Approximations*. Uppsala University, Department of Mathematics, Report No. 12, 16 pages.

[16] Forti, Mauro and Furio Honsell: 1983. *Set Theory with Free Construction Principles*. Annali Scuola Normale Superiore di Pisa, Classe di Scienze 10, 493–522.

[17] Jäger, Gerhard: 1993. Fixed points in Peano arithmetic with ordinals. *Annals of Pure and Applied Logic* 60: 119–132.

[18] Kreisel, Georg: 1960. Ordinal logics and the characterization of informal concepts of proof. In: *Proceedings of the 1958 International Congress of Mathematicians*, Edinburgh: 289–299.

[19] Kripke, Saul: 1975. Outline of a theory of truth. *The Journal of Philosophy* 72: 53–81.

[20] Lewis, David: 1996. *Convention, A Philosophical Study*. Cmabridge, MA: Harvard University Press.

[21] Lindström, Ingrid: 1989. A construction of non-well-founded sets within Martin-Löf type theory. *Journal of Symbolic Logic* 54: 57–64.

[22] Martin-Löf, Per: 1984. *Intuitionistic Type Theory*. Naples: Bibliopolis.

[23] Myhill, John: 1975. Constructive set theory. *Journal of Symbolic Logic* 40: 347–382.

[24] Rathjen, Michael: 1992. A proof–theoretic characterization of the primitive recursive set functions. *Journal of Symbolic Logic* 57: 954–969.

[25] Rathjen, Michael: 2000. The strength of Martin-Löf type theory with a superuniverse. Part I. *Archive for Mathematical Logic* 39: 1–39.

[26] Rathjen, Michael: 2001. Kripke-Platek set theory and the anti-foundation axiom. *Mathematical Logic Quarterly* 47: 435–440.

[27] Rathjen, Michael: 2003. The anti-foundation axiom in constructive set theories. In: G. Mints and R. Muskens (eds.), *Games, Logic, and Constructive Sets*, Stanford: CSLI Publications, 87–108.

[28] Russell, Bertrand: 1903. *Principles of Mathematics*. Cambridge: Cambridge University Press.

[29] Russell, Bertrand: 1908. Mathematical logic as based on the theory of types. *American Journal of Mathematics* 30: 222–262.

[30] Schütte, Kurt: 1964. Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik. *Archiv für Mathematische Logik und Grundlagenforschung* 67: 45–60.

[31] Schütte, Kurt: 1965. Predicative well-orderings. In: J. Crossley and M. Dummett (eds.), *Formal Systems and Recursive Functions*, Amsterdam: North-Holland, 176–184.

[32] Schütte, Kurt: 1977. *Proof Theory*. Berlin: Springer.

[33] Simpson, Stephen: 1999. *Subsystems of Second Order Arithmetic*. Berlin: Springer.

[34] Weyl, Hermann: 1949. *Philosophy of Mathematics and Natural Sciences*. Princeton: Princeton University Press.

School of Mathematics
University of Leeds
Leeds LS2 9JT
UK

E-mail: michrathjen@snafu.de

Russell's Paradox and Diagonalization in a Constructive Context

John L. Bell

Abstract. One of the most familiar uses of the Russell paradox , or, at least, of the idea underlying it, is in proving Cantor's theorem that the cardinality of any set is strictly less than that of its power set. The other method of proving Cantor's theorem—employed by Cantor himself in showing that the set of real numbers is uncountable—is that of diagonalization. Typically, diagonalization arguments are used to show that function spaces are "large" in a suitable sense. Classically, these two methods are equivalent. But constructively they are not: while the argument for Russell's paradox is perfectly constructive, (i.e., employs intuitionistically acceptable principles of logic) the method of diagonalization fails to be so. I describe the ways in which these two methods diverge in a constructive setting.

One of the most familiar uses of the Russell paradox, or, at least, of the idea underlying it, is in proving Cantor's theorem that, for any set E, the cardinality of E is strictly less than that of its power set $\mathcal{P}E$. This, as we all know, boils down to showing that there can be no surjection $E \twoheadrightarrow \mathcal{P}E$. To establish this it is enough to show that, for any map $f\colon E \to \mathcal{P}E$, the assertion

$$\forall X \in \mathcal{P}E\ \exists x \in E.\, X = f(x) \tag{1}$$

leads to a contradiction. Now from a *constructive* standpoint the argument of Russell's paradox establishes *more* than just the negation of (1), since it produces an explicit set R for which it can be proved that

$$\neg\exists x \in E.\, R = f(x)$$

namely the familiar "Russell set"

$$R = \{x \in E : x \notin f(x)\}.$$

This is, of course, because assuming $R = f(e)$ for some $e \in E$ leads instantly to the contradiction $e \in f(e) \Leftrightarrow e \notin f(e)$. For $U \subseteq E$, a similar argument, replacing R above by $R \cap U$, shows that there can be no surjection $U \twoheadrightarrow \mathcal{P}E$. If we agree to say that a set B is *surjective* with a set A provided that there is a surjection $A \twoheadrightarrow B$, then this may be put: for no set E is $\mathcal{P}E$ surjective with a subset of E.

These arguments are constructively valid in that they employ only constructively, or intuitionistically, acceptable principles of logic.

Equally, the (classically equivalent, but not automatically constructively equivalent) form of Cantor's theorem that, for any set E there is no injection $\mathcal{P}E \rightarrowtail E$ can also be given a constructive proof using the idea of Russell's paradox. In fact we can prove more, to wit, that for any set E there can be no injection of $\mathcal{P}E$ into a set *surjective* with E. For suppose given a surjection $f\colon E \twoheadrightarrow A$ and an injection $m\colon \mathcal{P}E \rightarrowtail A$. Define

$$B = \{x \in E : \exists X \in \mathcal{P}E.\, m(X) = f(x) \wedge x \notin X\}.$$

Since f is surjective there is $b \in E$ for which $f(b) = m(B)$. Then we have

$$\begin{aligned} b \in B &\iff \exists X.\, m(X) = f(b) \wedge b \notin X \\ &\iff \exists X.\, m(X) = m(B) \wedge b \notin X \\ &\iff \exists X.\, X = B \wedge b \notin X \\ &\iff b \notin B, \end{aligned}$$

and we have our contradiction.

As pointed out by George Boolos in [2], one can, classically, produce explicit counterexamples to the injectivity of a given map $m\colon \mathcal{P}E \to E$, that is, subsets X and Y of E for which $X \neq Y$ and $m(X) = m(Y)$. To do this it suffices to define a partial right inverse r of m such that, for $M = \mathrm{dom}(r)$, $m(M) \in M$ and $\forall x \in M.\ x \notin r(x)$. For then, writing $m(M) = a$, and $X = r(a)$, we have $a \notin X$, whence $X \neq M$, and $m(X) = m(r(a)) = a = m(M)$. Using an idea that goes back to Zermelo, Boolos obtains M as the field of the largest partial well-ordering $<$ of E such that $m(\{y : y < x\}) = x$ for all $x \in M$, and defines r by $r(x) = \{y : y < x\}$. The presence of well-orderings in this argument makes it highly nonconstructive; I do not know whether the existence of such an r and M can be established constructively.

Classically, the power set $\mathcal{P}E$ is naturally bijective with 2^E, the set of all maps $E \to 2 = \{0, 1\}$. Constructively, this is no longer the case: here, in general, $\mathcal{P}E \cong \Omega^E$, where Ω is the object of truth values or propositions, which is only $\cong 2$ when the law of excluded middle is assumed. In fact, constructively, 2^E is isomorphic, not to $\mathcal{P}E$, but to its Boolean sublattice $\mathcal{C}E$ consisting of all *complemented* (or *detachable*) subsets of E (a subset U of E is said to be complemented if $\forall x \in E.\ x \in U \vee x \notin U$). What happens when we replace $\mathcal{P}E$ by $\mathcal{C}E$ in the above arguments? Classically, of course, it makes no difference, but do the "Russell paradox" arguments survive the transition to constructivity?

Well, if one takes the first argument, showing that there can be no surjection $f\colon E \twoheadrightarrow \mathcal{P}E$, one finds that, when $\mathcal{P}E$ is replaced by $\mathcal{C}E$, the set $R \notin \mathrm{range}(f)$ is itself complemented and the argument goes through, proving constructively that there can be no surjection $E \twoheadrightarrow \mathcal{C}E$. But the second argument, with $\mathcal{P}E$ replaced by $\mathcal{C}E$ (and E replaced by a subset U of E) only goes through constructively when U is itself complemented. And as for the third argument to go through constructively once $\mathcal{P}E$ is replaced by $\mathcal{C}E$, it is necessary to show that the set B defined there is complemented, and, as we shall see, this cannot in general be done. The failure of these two latter

arguments in a constructive context can be easily demonstrated by considering a model $\mathfrak{M}$ of smooth infinitesimal analysis, see, e.g., [1]. In $\mathfrak{M}$ the real line $\mathbb{R}$ has just the two detachable subsets $\emptyset, \mathbb{R}$, that is, $\mathcal{C}\mathbb{R}$ has just two elements. A *fortiori* $\mathcal{C}\mathbb{R}$ is injectible into $\mathbb{R}$, showing that the third argument fails constructively. The fact that there is no surjection $\mathbb{R} \to \mathcal{C}\mathbb{R}$ corresponds simply to the fact that, since $\mathbb{R}$ is connected, there are no continuous nonconstant maps $\mathbb{R} \to 2$ (all maps in $\mathfrak{M}$ being smooth, and certainly continuous). But $\mathcal{C}\mathbb{R}$ is trivially surjective with the subset $\{0, 1\}$ of $\mathbb{R}$, refuting the second argument—of course, in $\mathfrak{M}$, $\{0, 1\}$ is not a complemented subset of $\mathbb{R}$! At the end of the paper we supply a quite different constructive example of a set E for which $\mathcal{C}E$ is surjective with a subset of E.

Let us return once again to the argument that there is no surjection $E \twoheadrightarrow \mathcal{P}E$. Classically, we may replace $\mathcal{P}E$ by the isomorphic object 2^E. In that case the use of Russell's paradox is transformed into an application of *diagonalization*, the technique Cantor used to prove that the set of real numbers has strictly larger cardinality than the set of natural numbers. Indeed, if $\phi\colon E \to 2^E$ is the map canonically associated with the given map $f\colon E \to \mathcal{P}E$ via characteristic functions (i.e., defined by $\phi(x)(y) = 1 \Leftrightarrow y \in f(x)$) then the "Russell" set $R \in \mathcal{P}E$ outside the range of f corresponds precisely to the map $r\colon E \to 2$ outside the range of ϕ and defined by "diagonalization":

$$r(x) = \begin{cases} 0 & \text{if } \phi(x)(x) = 1 \\ 1 & \text{if } \phi(x)(x) = 0. \end{cases}$$

This argument is perfectly constructive and parallels that given above for the nonexistence of a surjection $E \twoheadrightarrow \mathcal{C}E$.

Diagonalization appears in one of its most familiar guises in the well-known proof that there can be no surjection of the set $\mathbb{N}$ of natural numbers onto the set $\mathbb{N}^{\mathbb{N}}$ of all maps $\mathbb{N} \to \mathbb{N}$, that, in a word, $\mathbb{N}^{\mathbb{N}}$ is *uncountable*. Here one is given a map $\phi\colon \mathbb{N} \to \mathbb{N}^{\mathbb{N}}$; then the map $f\colon \mathbb{N} \to \mathbb{N}$ defined by the prescription

$$f(n) = \begin{cases} 0 & \text{if } \phi(n)(n) \neq 0 \\ 1 & \text{if } \phi(n)(n) = 0. \end{cases}$$

is clearly outside the range of φ. This instance of diagonalization, although not identical with the previous one, also appears very similar to Russell's paradox.

Looked at constructively, this proof depends crucially on the *decidability* of $\mathbb{N}$, i.e., the truth of the assertion

$$\forall m \in \mathbb{N}\ \forall n \in \mathbb{N}.\ m = n \vee m \neq n.$$

Since $\mathbb{N}$ is decidable from a constructive standpoint, the argument is constructively valid, so that $\mathbb{N}^{\mathbb{N}}$ is constructively uncountable. (More generally, the same argument shows that if X is any decidable set with at least two distinct elements, X^X cannot be surjective with X.)

Let us call a set *subcountable* if it is surjective with a subset of $\mathbb{N}$. Classically, it follows trivially from the fact that $\mathbb{N}^{\mathbb{N}}$ is uncountable that it also fails to be *subcount-*

able, that is, $\mathbb{N}^{\mathbb{N}}$ is not surjective with a subset of $\mathbb{N}$. For given $\phi\colon U \to \mathbb{N}^{\mathbb{N}}$; the map $f\colon \mathbb{N} \to \mathbb{N}$ defined by the new prescription (still resembling Russell's paradox)

$$f(n) = \begin{cases} 0 & \text{if } n \in U \ \& \ \phi(n)(n) \neq 0 \\ 1 & \text{if } n \in U \ \& \ \phi(n)(n) = 0 \\ 1 & \text{if } n \notin U \end{cases}$$

is again clearly outside the range of ϕ. Now this argument only goes through constructively when U is a detachable subset of $\mathbb{N}$, so that the most "diagonalization" shows, constructively, is that $\mathbb{N}^{\mathbb{N}}$ is not "detachably" *subcountable*. But *prima facie* nothing prevents $\mathbb{N}^{\mathbb{N}}$ from being "nondetachably" subcountable; for if, given $\phi\colon U \to \mathbb{N}^{\mathbb{N}}$ we repeat the original prescription, the map f so obtained is defined just on U, not on the whole of $\mathbb{N}$, and the argument collapses. So, while it still follows from Russell's paradox that $\mathcal{P}\mathbb{N}$ cannot be constructively subcountable, diagonalization (as well as Russell's paradox) fails to establish the corresponding fact for $\mathbb{N}^{\mathbb{N}}$.

In fact models of constructive mathematics have been produced in which $\mathbb{N}^{\mathbb{N}}$ *is actually subcountable*. Such is the case, notably, in the *effective topos* **Eff**, see, e.g., chap. 23 of [3]. In **Eff**, $\mathcal{P}\mathbb{N}$ and $\mathbb{N}^{\mathbb{N}}$ effectively(!) part company, making Russell's paradox and diagonalization the more easily distinguished. Russell's paradox continues to yield the non-subcountability of $\mathcal{P}\mathbb{N}$, which accordingly remains "large". But diagonalization, while continuing to yield the uncountability of $\mathbb{N}^{\mathbb{N}}$, fails to prevent it from being subcountable in **Eff**, and so from being in some sense "small" there. The reason for this is that, in **Eff**, $\mathbb{N}^{\mathbb{N}}$ consists, not of arbitrary maps $\mathbb{N} \to \mathbb{N}$, but just of the *recursive* ones. The subset $U \subseteq \mathbb{N}$ establishing the subcountability of $\mathbb{N}^{\mathbb{N}}$ is the set of codes of total recursive functions; since, in **Eff**, the complemented subsets of $\mathbb{N}$ are just the recursive subsets, the fact that U cannot be complemented corresponds to the fact that the set of codes of total recursive functions is not itself recursive. The subcountability of $\mathbb{N}^{\mathbb{N}}$ immediately implies that of its subset 2^N, and hence also of the latter's isomorph $\mathcal{C}\mathbb{N}$, showing anew the failure, in a constructive context, of the argument that there can be no set E for which $\mathcal{C}E$ is surjective with a subset of E.

The "divergence" in **Eff** between diagonalization and Russell's paradox can be further pointed up by observing that **Eff** contains *nonsingleton* objects C for which the object C^C of self-maps is actually *isomorphic* to C. Such objects cannot be classical sets, because clearly the only such sets satisfying this condition are singletons. For a classical set C, the condition of being a singleton is equivalent to the condition that there be no injection of 2—the object of truth values in classical set theory—into C. So, in classical set theory, the condition that $C^C \cong C$ implies that there is no injection of Ω into C. In fact Russell's paradox shows that *this* implication continues to hold in the constructive setting. For suppose that i is an isomorphism (even just an injection) of C^C with C and that m is an injection of Ω into C. Then the map

$$X \mapsto i(\{\langle x, m(x \in X)\rangle : x \in C\})$$

is easily seen an injection of $\mathcal{P}C$ into C, which by the Russell's paradox argument above, is impossible.

To summarize: in a constructive context, diagonalization does not fail to prevent the possible presence of an object which is isomorphic to its object of self-maps and yet is not a singleton. But Russell's paradox does preclude the existence of an injection of the object of truth values into any such object C: such a map would yield an injection into C of its power object $\mathcal{P}C$, which by Russell's paradox is too large to be so injectible.

To conclude. We have seen that, in certain constructive contexts, diagonalization may fail to ensure that function spaces are "relatively large". By contrast, Russell's paradox—at least, as properly applied to power sets—retains its potency even in constructive environments, ensuring that power sets, or objects, retain their "size". It seems fitting, therefore, to claim for Russell's paradox a universal applicability which must at the same time be denied diagonalization.

References

[1] Bell, John L.: 1998. *A Primer of Infinitesimal Analysis*. Cambridge: Cambridge University Press.

[2] Boolos, George: 1997. Constructing cantorian counterexamples. *Journal of Philosophical Logic* 26: 237–239.

[3] McLarty, Colin: 1992. *Elementary Categories, Elementary Toposes*. New York: Oxford University Press.

Department of Philosophy
Talbot College
University of Western Ontario
London, Ontario
Canada N6A 3K7

E-mail: jbell@uwo.ca

Constructive Solutions of Continuous Equations

Peter Schuster and Helmut Schwichtenberg

Abstract. In this paper we modify some seminal notions from constructive analysis by providing witnesses for (strictly) positive quantifiers occurring in their definitions. For instance, by a real number we understand nothing but a modulated Cauchy sequence of rationals. This approach allows us to avoid many of the putatively necessary appeals to countable choice. Accordingly, some basic machinery of constructive analysis is made explicit; this includes choice-free proofs of the sequential and order completeness of the real numbers, and of the approximate intermediate value theorem. Moreover, we unwind the Kneser proof of the fundamental theorem of algebra by giving a (rather detailed) choice-free proof.

1. Introduction

Since its conception in Bishop's seminal monograph [3], constructive analysis has attracted the attention of many researchers. In particular, it poses a challenge to design logical systems which allow for a smooth formalization. There are obvious reasons why one wants to be clear about the possibility of such formalizations, among them the prospect of using formal proofs of existential theorems as a means to develop correct programs, in this case, involving exact real numbers.

For such purposes it is important that the system does not include (countable or dependent) choice axioms. The aim of the present paper is to sketch the development of elementary constructive analysis on this basis, up to and including the fundamental theorem of algebra. The key to such a choice-free treatment is to provide witnesses of (strictly) positive existential quantifiers in all definitions.

2. Elementary Constructive Analysis

We understand real numbers as modulated Cauchy sequences of rationals.[1] So a *real number,* or simply *real*, is nothing but a sequence $(a_m)_{m\in\mathbb{N}}$ of rational numbers together with a modulus $M\colon \mathbb{Q}^+ \to \mathbb{N}$ such that $|a_m - a_n| \le \varepsilon$ for $m, n \ge M(\varepsilon)$. Let

[1] Inasmuch as we do not focus on fixed moduli, we deviate from [2, 3, 21]. An approach similar to the present one was taken in [22, 16].

us stress that throughout this paper we strictly identify the notions "real number" and "modulated Cauchy sequence of rationals" with one another, even though we sometimes say that an instance of the latter represents, gives, or defines the corresponding instance of the former. As usual, we denote the set[2] of real numbers by $\mathbb{R}$.

Every rational q is tacitly understood as the real represented by the constant sequence $a_m = q$ with the constant modulus $M(\varepsilon) = 1$. For any rational q with $|q| < 1$, a less trivial representation of a real (in fact, of the rational $\frac{1}{1-q}$) is that given by the partial sums $a_n = \sum_{i=0}^{n} q^i$ of the geometric series: as $|a_m - a_n| \le \frac{2}{|1-q|}|q|^{n+1}$ whenever $m \ge n$, an appropriate modulus is $M(\varepsilon) = \min\{\, n \in \mathbb{N} \mid |q|^{n+1} \le \frac{|1-q|}{2}\varepsilon \,\}$.

A real a is *nonnegative* (written as $a \in \mathbb{R}^{0+}$) if $-\varepsilon \le a_{M(\varepsilon)}$ for all rationals $\varepsilon > 0$, and *positive* (written as $a \in \mathbb{R}^{+}$) if $\varepsilon \le a_{M(\frac{\varepsilon}{2})}$ for some rational $\varepsilon > 0$. We write $a \in_\varepsilon \mathbb{R}^+$ whenever the rational $\varepsilon > 0$ is such a witness of $a \in \mathbb{R}^+$. So if $a \in_\varepsilon \mathbb{R}^+$, then $a_m \ge \frac{\varepsilon}{2}$ for all $m \ge M(\frac{\varepsilon}{2})$.

Given real numbers a, b represented by sequences of rationals $(a_m)_{m\in\mathbb{N}}$, $(b_m)_{m\in\mathbb{N}}$ with moduli M, N, we define each c of the list $a+b$, $-a$, $|a|$, $a\cdot b$, and $\frac{1}{a}$ (the latter only provided that $|a| \in_\eta \mathbb{R}^+$) as represented by the respective sequence (c_m) of rationals with modulus L:

c	c_m	$L(\varepsilon)$
$a + b$	$a_m + b_m$	$\max\{M(\frac{\varepsilon}{2}), N(\frac{\varepsilon}{2})\}$
$-a$	$-a_m$	$M(\varepsilon)$
$\lvert a\rvert$	$\lvert a_m\rvert$	$M(\varepsilon)$
$a \cdot b$	$a_m \cdot b_m$	$\max\{M(\frac{\varepsilon}{2K_{\lvert b\rvert}}), N(\frac{\varepsilon}{2K_{\lvert a\rvert}})\}$
$\frac{1}{a}$ for $\lvert a\rvert \in_\eta \mathbb{R}^+$	$\begin{cases} \frac{1}{a_m} & \text{if } a_m \neq 0 \\ 0 & \text{if } a_m = 0 \end{cases}$	$\max\{M(\frac{\eta}{2}), N(\frac{\varepsilon\eta^2}{4})\}$

where the rational number $K_a = \max\{\, a_n \mid n \le M(1) \,\} + 1$ is such that $a_m \le K_a$ for all $m \in \mathbb{N}$. Needless to say, $a - b$ and $\frac{a}{b}$ (for $b \in \mathbb{R}^+$) stand for $a + (-b)$ and $a \cdot \frac{1}{b}$, respectively.

We write $a \le b$ for $b - a \in \mathbb{R}^{0+}$ and $a < b$ for $b - a \in \mathbb{R}^+$. Unwinding the definitions yields that $a \le b$ means that for every rational $\eta > 0$ there exists $P_\eta \in \mathbb{N}$ such that $a_m \le b_m + \eta$ for all $m \ge P_\eta$. Furthermore, $a < b$ is a shorthand way of expressing the presence of a rational $\eta > 0$ and a positive integer Q with $a_m + \eta \le b_m$ for all $m \ge Q$; we then write $a <_{\eta,Q} b$ (or simply $a <_\eta b$ if Q is not needed) whenever we want to call these witnesses.

Constructively, equality $a = b$ and inequality $a \neq b$ between real numbers are to be defined as $a \ge b \wedge a \le b$ and $a < b \vee a > b$, respectively. Some routine proofs show that the arithmetic of real numbers respects equality. Moreover, $a \le b$ could equivalently be defined as $\neg(a > b)$, and $a = b$ as $\neg(a \neq b)$, whereas $a < b$ and $a \neq b$

[2] As we hesitate to commit ourselves to any set theory whatsoever, we wish to understand each set as identified with the defining property of its elements.

are constructively stronger than $\neg(a \geq b)$ and $\neg(a = b)$, respectively.[3] Intervals are then defined in the familiar way; so $[a, b]$ is defined as the set $\{\, c \in \mathbb{R} \mid a \leq c \leq b \,\}$.

Except for the completeness properties, which we shall subsequently discuss, it is routine to verify that the reals as defined above possess all the features of the reals that are commonly considered to be constructive [4]. Constructively, in particular, we cannot compare two reals, but we can compare each real with every nontrivial interval, as in the following *approximate splitting principle*.

Lemma 1. *For all real numbers a, b, c, if $a < b$, then either $a < c$ or $c < b$.*

Proof. Let M, N, L denote the moduli of a, b, c, respectively, and assume that $a <_{7\eta, P} b$, i.e.,

$$a_m + 3\eta < a_m + 4\eta \leq b_m - 3\eta$$

for every $m \geq P$. For $Q = \max\{M(\eta), N(\eta), L(\eta), P\}$, we have either $c_Q < a_Q + 4\eta$ or else $a_Q + 4\eta \leq c_Q$, but not both. In the former case, $c_Q < b_Q - 3\eta$, and thus

$$c_m + \eta \leq c_Q + 2\eta < b_Q - \eta \leq b_m$$

for every $m \geq Q$, which means that $c <_{\eta, Q} b$. In the latter case, $a_Q + 3\eta < c_Q$, and thus

$$a_m + \eta \leq a_Q + 2\eta < c_Q - \eta \leq c_m$$

for every $m \geq Q$, which is to say that $a <_{\eta, Q} c$. □

It is noteworthy that although $a < c$ and $c < b$ may well happen simultaneously, the proof of Lemma 1 contains a routine that automatically pins down one of the alternatives in every such case. This virtue, which of course stands and falls with the presence of the moduli for a, b, c and the witness for $a < b$, is of particular relevance whenever the approximate splitting principle is applied infinitely often, so in the proof of Theorem 12 below. By employing that routine, we can avoid the usual invocation of dependent choice to successively construct a sequence of real numbers.

An immediate consequence of Lemma 1 is that $b \leq c$ means that $a < b$ implies $a < c$ for all reals c.

For $1 \leq j \leq n$ let a_j be a real number given by the sequence of rationals $(a_{jm})_{m \in \mathbb{N}}$ with modulus M_j. Then the *maximum* $\max\{\, a_j \mid j \leq n \,\}$ is the real number represented by the sequence of rationals $\big(\max\{\, a_{jm} \mid j \leq n \,\}\big)_{m \in \mathbb{N}}$ with modulus $\max\{\, M_j \mid j \leq n \,\}$. Constructively, one cannot pick an index i for which $a_i = \max\{\, a_j \mid j \leq n \,\}$, but for each $\varepsilon > 0$ we can find—as follows—an index i such that a_i is maximal *up to* ε among all the a_j.

Lemma 2. *For all real numbers $a_1, \ldots, a_n$ and each rational $\varepsilon > 0$ we can find an $i \leq n$ so that $a_j \leq a_i + \varepsilon$ for every $j \leq n$.*

[3] This is clear in view of the existential and universal character of the predicates $\in \mathbb{R}^+$ and $\in \mathbb{R}^{0+}$, respectively.

Proof. For each $j \le n$ let a_j be represented by the sequence of rationals $(a_{jm})_{m\in\mathbb{N}}$ with modulus M_j. Given a rational $\varepsilon > 0$, set

$$M = \max\left\{ M_j\left(\frac{\varepsilon}{2}\right) \mid j \le n \right\}, \quad b = \max\{\, a_{jM} \mid j \le n \,\},$$

and

$$i = \min\{\, k \le n \mid a_{kM} = b \,\}.$$

For each $j \le n$ we then have

$$a_{jm} \le a_{jM} + \frac{\varepsilon}{2} \le a_{iM} + \frac{\varepsilon}{2} \le a_{im} + \varepsilon$$

for every $m \ge M$, so that $a_j \le a_i + \varepsilon$ as was required. □

As for Lemma 1 above, we wish to stress that by virtue of the presence of moduli the algorithm given in the proof of Lemma 2 always picks a definite index i even when there are several fulfilling its specification. This feature again allows us to do proofs without countable choice.

Let us now have a look at the completeness properties characteristic of the real numbers. We first consider sequential completeness. A sequence $(a_m)_{m\in\mathbb{N}}$ of reals is a *Cauchy sequence* with modulus $M\colon \mathbb{Q}^+ \to \mathbb{N}$ whenever $|a_m - a_n| \le \varepsilon$ for $m, n \ge M(\varepsilon)$, and *converges* with modulus $M\colon \mathbb{Q}^+ \to \mathbb{N}$ to a real b, its *limit*, whenever $|a_m - b| \le \varepsilon$ for $m \ge M(\varepsilon)$.

Lemma 3. *Every modulated Cauchy sequence of rationals converges with the same modulus to the real number it represents.*

Proof. Let $(a_m)_{m\in\mathbb{N}}$ be a Cauchy sequence of rationals with modulus M, and let b be the real number it represents. For each rational $\varepsilon > 0$ and every $m \ge M(\varepsilon)$, we have to prove that $|a_m - b| \le \varepsilon$, which is to say that $c_m \ge 0$ for $c_m = \varepsilon - |a_m - b|$. Now the real c_m is represented by the sequence $\big(\varepsilon - |a_m - a_k|\big)_{k\in\mathbb{N}}$ of rationals, which again has modulus M: for every rational $\eta > 0$ and all $k, l \ge M(\eta)$, we have

$$|c_k - c_l| = \big||a_m - a_k| - |a_m - a_l|\big| \le \big|(a_m - a_k) - (a_m - a_l)\big| = |a_k - a_l| \le \eta.$$

Therefore $c_m \ge 0$ is equivalent to $-\eta \le \varepsilon - |a_m - a_{M(\eta)}|$ for all rationals $\eta > 0$. This follows from

$$|a_m - a_{M(\eta)}| \le |a_m - a_{\max\{M(\varepsilon),M(\eta)\}}| + |a_{\max\{M(\varepsilon),M(\eta)\}} - a_{M(\eta)}| \le \varepsilon + \eta\,,$$

which holds by virtue of the inequality $m \ge M(\varepsilon)$. □

By the triangle inequality, every convergent sequence of reals with modulus M is a Cauchy sequence with modulus $\varepsilon \mapsto M(\frac{\varepsilon}{2})$. Writing $\lceil\varepsilon\rceil$ for the least integer $\ge \varepsilon$ whenever $\varepsilon > 0$ is rational, we now prove the reverse implication.[4]

[4] This stems from unpublished lecture notes of the second author.

Theorem 4 (Sequential Completeness). *For every modulated Cauchy sequence of reals we can find a real number to which it converges with modulus.*

Proof. Let $(a_m)_{m\in\mathbb{N}}$ be the Cauchy sequence of reals with modulus M; for every $m \in \mathbb{N}$, let a_m be represented by a Cauchy sequence $(a_{mk})_{k\in\mathbb{N}}$ of rationals with modulus M_m. Note first that, for each $m \in \mathbb{N}$ and every rational $\eta > 0$, by Lemma 3 we have $|a_m - a_{ml}| \leq \frac{1}{m}$ for all $l \geq M_m(\frac{1}{m})$. Next, set

$$b_m = a_{m M_m\left(\frac{1}{m}\right)}$$

for every $m \in \mathbb{N}$, so that $|a_m - b_m| \leq \frac{1}{m}$ for all $m \in \mathbb{N}$ by the particular case $l = M_m\left(\frac{1}{m}\right)$ of the foregoing consideration. Then

$$|b_m - b_n| \leq |b_m - a_m| + |a_m - a_n| + |a_n - b_n| \leq \frac{1}{m} + \frac{\varepsilon}{2} + \frac{1}{n} \leq \varepsilon$$

for all $m, n \geq N(\varepsilon) = \max\left\{M(\frac{\varepsilon}{2}), \lceil\frac{4}{\varepsilon}\rceil\right\}$, which is to say that (b_m) is a Cauchy sequence with modulus N, and therefore represents a real number b. Moreover,

$$|a_m - b| \leq |a_m - b_m| + |b_m - b| \leq \frac{1}{m} + \frac{\varepsilon}{2} \leq \varepsilon$$

for all $m \geq N(\frac{\varepsilon}{2}) = \max\left\{M(\frac{\varepsilon}{4}), \lceil\frac{8}{\varepsilon}\rceil\right\}$; in other words: (a_m) converges with b with modulus $\varepsilon \mapsto N(\frac{\varepsilon}{2})$. □

This proof happens not to need countable choice but requires a certain conceptual choice made in advance: that each real number is meant to be nothing but a modulated Cauchy sequence of rationals. Given this presupposition, we have no reservations from the choice-free perspective that a single sequence of reals is understood as a double sequence of rationals.[5]

On the other hand, countable choice would become indispensable for the proof of sequential completeness as soon as the concept of reals were relaxed either to Cauchy sequences of rationals without moduli or else to equivalence classes, modulo =, of modulated Cauchy sequences of rationals. To start our proof of Theorem 4 in the former case, one must choose, for every m, the integer l with $|a_m - a_{ml}| \leq \frac{1}{m}$, which in the presence of the modulus M_m of a_m we can simply define by setting $l = M_m(\frac{1}{m})$. In the latter case, one must choose first, again for every m, a representative $(a_{mk})_{k\in\mathbb{N}}$ of the real a_m, which we understand as already consisting in nothing but such a sequence of rationals.

[5] When one wants to read this proof as a completeness proof for the sequential completion of an arbitrary metric space, one again has to view sequences in this completion—the space of (modulated) Cauchy sequences in the original space—as double sequences in the latter. Such an interpretation may, however, look somewhat strange in cases which are more complex than that of the completion of the rationals to the reals.

For this observation, and for many others on the role of countable choice in constructive mathematics, the first author is indebted to Fred Richman, for whose position we refer to [15].

Following [2, 3, 5] in the style of [19], let us now consider order completeness. To this end, let S be a set of real numbers. A real b is an *upper bound* of S, or S is *bounded from above* by b, whenever $s \leq b$ for all $s \in S$, and a *least upper bound* or *supremum* $\sup S$ of S if, in addition, for every real a with $a < b$ there is an $r \in S$ with $r > a$.[6]

If b is a least upper bound and a an upper bound of S, then $a < b$ is impossible, so that $b \leq a$; whence up to equality there is at most one supremum of S. If S has a supremum, then S is inhabited and bounded from above, but even when S has the two latter properties, the former is not granted constructively. A simple exception is the maximum of finitely many reals (see above), which of course is the supremum thereof.

A more relevant sufficient condition for the constructive existence of the supremum of a general set S of real numbers (being inhabited and bounded from above) is that S be *order located from above*[7], which is to say that

$$\Sigma_S(a,b): \quad \text{either } s \leq b \text{ for all } s \in S \text{ or } a < r \text{ for some } r \in S$$

holds for all pairs a, b of real numbers with $a < b$. To get at the sufficiency of these three conditions, it is useful to consider

$$\Pi_S(a,b): \quad \text{both } s \leq b \text{ for all } s \in S \text{ and } a < r \text{ for some } r \in S$$

as a property of any pair a, b of real numbers with $a < b$. Note that $\Sigma_S(a,b)$ and $\Pi_S(a,b)$ are the disjunction and conjunction, respectively, of the same pair of propositions.

Theorem 5 (Order Completeness). *Let $S \subset \mathbb{R}$ be inhabited, bounded from above, and order located from above. Then S has a least upper bound.*

Proof. We may assume that we have $a, b \in \mathbb{Q}$ with $a < b$ such that $\Pi_S(a,b)$. We construct $c, d : \mathbb{N} \to \mathbb{Q}$ such that for all m

$$a = c_0 \leq c_1 \leq \cdots \leq c_m < d_m \leq \cdots \leq d_1 \leq d_0 = b, \tag{1}$$

$$\Pi_S(c_m, d_m), \tag{2}$$

$$d_m - c_m \leq \left(\frac{2}{3}\right)^m (b-a). \tag{3}$$

Let $c_0, \ldots, c_m$ and $d_0, \ldots, d_m$ be already constructed such that (3) holds. Let $x = c_m + \frac{1}{3}(d_m - c_m)$ and $y = c_m + \frac{2}{3}(d_m - c_m)$. Since S is order located from above, either $s \leq y$ for all $s \in S$ or else $x < r$ for some $r \in S$. In the first case let $c_{m+1} := c_m$ and $d_{m+1} := y$, and in the second case let $c_{m+1} := x$ and $d_{m+1} := d_m$. Then clearly $\Pi_S(c_{m+1}, d_{m+1})$, and also (1) and (3) continue to hold for $m+1$. We claim that the

[6] This is a positive way of saying that every real a with $a < b$, however close it may be to b, is not an upper bound of S.

[7] This name was coined in [8]. The same property of S has been called "lower located" by Erik Palmgren [11] and (if, in addition, S is bounded above) "located above" by Peter Aczel and Michael Rathjen [1].

real numbers $c = d$ given by the modulated Cauchy sequence of rationals (c_m) and (d_m) is the least upper bound of S. Indeed, it is an upper bound because for $\varepsilon > 0$ we have $d \leq d_m < d + \varepsilon$ for some m and d_m is an upper bound of S, and similarly $c - \varepsilon < c_m \leq c$ for some m and $c_m < r$ for some $r \in S$. □

One might think that dependent choice has been used in this proof: in the case distinction "either $s \leq y$ for all $s \in S$ or else $x < r$ for some $r \in S$" the cases may well overlap. However, reference to the algorithm provided by the assumption that S is order located from above makes this choice superfluous.

Another important condition sufficient for the existence of a supremum is that S is a *totally bounded* set of reals. Let us consider this concept next.

To start with, let us call an arbitrary set *finite* if there is a mapping from $\{1, \ldots, n\}$ onto that set with $n \geq 1$; in particular, finite sets are inhabited.[8] A set S of real numbers is *totally bounded* whenever for every (rational) $\varepsilon > 0$ there is an *ε-approximation* of S: that is, a finite subset T of S such that each element of S is within ε of an element of T. Totally bounded sets of real numbers are obviously inhabited and bounded.

Note that the definition of total boundedness involves two witnessing objects: the sequence $(T_m)_{m \geq 1}$ of finite subsets of S so that each T_m is a $\frac{1}{m}$-approximation of S, and also for every m a function assigning to each element of S an element of T_m that lies within $\frac{1}{m}$ of the former. Notice also that the latter function, having a finite range, is necessarily non-continuous.

Clearly every totally bounded set of reals is order located from above; therefore we immediately obtain from Theorem 5:

Corollary 6. *Every totally bounded set of real numbers has a supremum.*

As a digression, let us recall the standard application of Corollary 6.

Proposition 7. *Let $D \subset \mathbb{R}$, and let $f\colon D \to \mathbb{R}$ be a uniformly continuous function. If the domain D of f is totally bounded, then so is the range $f(D)$ of f.*

Proof. For $\varepsilon > 0$, pick $\delta > 0$ so that, for all $x, y \in D$, if $|x - y| \leq \delta$, then $|f(x) - f(y)| \leq \varepsilon$. Now if T is a finite δ-approximation to D, then $f(T)$ is a finite ε-approximation to $f(D)$. □

Corollary 8. *If $f\colon D \to \mathbb{R}$ is a uniformly continuous function on a totally bounded subset D of $\mathbb{R}$, then the supremum of $f(D)$ exists.*

In particular, although one cannot expect a constructive method to locate a point in D at which f attains its supremum, for every $\varepsilon > 0$ there is an $x \in D$ so that

[8] In this respect we follow [2, 3] rather than [5, 10], where finite sets are allowed to be empty. For the sake of simplicity, we also use "finite" instead of the constructively more correct term "finitely enumerable" (due to Ray Mines) or "subfinite" (coined by Errett Bishop). Usually "finite" is reserved for the sets that are in a bijective correspondence to sets of the form $\{1, \ldots, n\}$, and which are necessarily discrete.

$\sup f(D) - \varepsilon < f(x)$. Compare this with Lemma 2, and note that the existence of the maximum of finitely many reals is a special case of Corollary 8.

Note that sequential completeness can be deduced from order completeness in a choice-free way [19][9]. However, when one literally transfers that deduction to the present context, it fails to automatically supply the required moduli. As a digression, let us thus provide a better suited argument, which still does not make any use of countable choice.

Reduction of sequential to order completeness. Let $(a_m)_{m\in\mathbb{N}}$ be a Cauchy sequence of real numbers, and define L as the set of all the $p \in \mathbb{Q}$ for which there is an $N \in \mathbb{N}$ so that $p < a_m$ for all $m \geq N$. This L is clearly inhabited and bounded above; we first show that L is order located from above. To this end, let $r < s$ be rational numbers. For $\varepsilon = \frac{s-r}{3}$, we have

$$a_{M(\varepsilon)} - \varepsilon \leq a_m \leq a_{M(\varepsilon)} + \varepsilon \quad \text{for all } m \geq M(\varepsilon),$$

and $r + \varepsilon < s - \varepsilon$, so that either $r + \varepsilon < a_{M(\varepsilon)}$ or $a_{M(\varepsilon)} < s - \varepsilon$. In the former case, $r < q < a_{M(\varepsilon)} - \varepsilon$ for some $q \in \mathbb{Q}$, for which $q \in L$. In the latter case, $p < s$ for every $p \in L$.

By order completeness, L has a supremum b. To conclude the proof, we now verify that (a_m) converges to b with modulus $\varepsilon \mapsto M\left(\frac{\varepsilon}{2}\right)$. Given a rational $\varepsilon > 0$, pick $p \in L$ with $b - \frac{\varepsilon}{2} < p$, and $N \in \mathbb{N}$ with $p < a_m$ for all $m \geq N$. For $K = \max\left\{N, M\left(\frac{\varepsilon}{2}\right)\right\}$, observe first that $b - \frac{\varepsilon}{2} < a_k$ and thus $b - \varepsilon < a_m$ for all $m \geq M\left(\frac{\varepsilon}{2}\right)$. Assume next that $a_k > b + \frac{\varepsilon}{2}$. Then $a_k - \frac{\varepsilon}{2} > b$, so that $b < p < a_k - \frac{\varepsilon}{2}$ for some rational p, for which, in particular, $p < a_m$ for all $m \geq M\left(\frac{\varepsilon}{2}\right)$ and thus $p \in L$, a contradiction to $p > b$. Hence $a_k \leq b + \frac{\varepsilon}{2}$ and thus also $a_m \leq b + \varepsilon$ for all $m \geq M\left(\frac{\varepsilon}{2}\right)$. □

Needless to say, the approach to greatest lower bounds (alias infima) is completely analogous to that of least upper bounds (alias suprema).

We next supply the standard constructive versions of the *intermediate value theorem*. A function $f : [a, b] \to \mathbb{R}$ is *locally nonconstant* whenever if $a \leq a' < b' \leq b$ and d is an arbitrary real, then $f(c) \neq d$ for some real $c \in [a', b']$. Note that if f is continuous, then there is also a rational c with that property. Strictly monotonic functions are clearly locally nonconstant, and so are nonconstant real polynomials.

Proposition 9 (Intermediate Value Theorem for Locally Nonconstant Functions). *Let $a < b$ be real numbers. If $f : [a, b] \to \mathbb{R}$ is continuous with $f(a) < 0 < f(b)$ and locally nonconstant, then we can find $c \in [a, b]$ with $f(c) = 0$.*

Proof. We may assume that $a, b \in \mathbb{Q}$. We construct $c, d : \mathbb{N} \to \mathbb{Q}$ such that for all m

$$a = c_0 \leq c_1 \leq \cdots \leq c_m < d_m \leq \cdots \leq d_1 \leq d_0 = b, \tag{4}$$

$$f(c_m) < 0 < f(d_m), \tag{5}$$

[9] In addition, the deduction given in [4] can easily be reformulated without choice.

$$d_m - c_m \leq \left(\frac{2}{3}\right)^m (b-a). \tag{6}$$

Let $c_0, \ldots, c_m$ and $d_0, \ldots, d_m$ be already constructed such that (6) holds. Let $x = c_m + \frac{1}{3}(d_m - c_m)$ and $y = c_m + \frac{2}{3}(d_m - c_m)$. By assumption we have a rational $z \in [c_m, d_m]$ with $f(z) \neq 0$. In case $0 < f(z)$ let $c_{m+1} = c_m$ and $d_{m+1} = z$, and in case $f(z) < 0$ let $c_{m+1} = z$ and $d_{m+1} = d_m$. As f is continuous, $f(c) = 0 = f(d)$ for the real numbers $c = d$ given by the modulated Cauchy sequence of rationals (c_m) and (d_m). □

One might think that dependent choice has been used in this proof: whether $f(z) > 0$ or $f(z) < 0$ may well depend on the choice of z. However, this is not so, inasmuch as we can refer to the algorithm contained in the assumption that f is locally nonconstant.

In Section 3, however, we can do the proof of the fundamental theorem of algebra with only the following basic tool at hand. Notice that clearly every real polynomial is uniformly continuous on $[a, b]$.

Theorem 10 (Approximate Intermediate Value Theorem). *Let $a < b$ be real numbers. For every uniformly continuous function $f : [a, b] \to \mathbb{R}$ with $f(a) \leq 0 \leq f(b)$, and every $\varepsilon > 0$, we can find $c \in [a, b]$ such that $|f(c)| \leq \varepsilon$.*

Proof. In the sequel we repeatedly invoke Lemma 1. Let $\varepsilon > 0$, and compare $f(a)$ and $f(b)$ with $-\varepsilon < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < \varepsilon$, respectively. If $-\varepsilon < f(a)$ or $f(b) < \varepsilon$, then $|f(c)| < \varepsilon$ for $c = a$ or $c = b$; whence we may assume that

$$f(a) < -\frac{\varepsilon}{2} \quad \text{and} \quad \frac{\varepsilon}{2} < f(b).$$

Now pick $\delta > 0$ so that, for all $x, y \in [a, b]$, if $|x-y| \leq \delta$, then $|f(x)-f(y)| \leq \varepsilon$, and divide $[a, b]$ into $a = a_0 < a_1 < \cdots < a_m = b$ such that $|a_{i-1} - a_i| \leq \delta$. Compare every $f(a_i)$ with $-\frac{\varepsilon}{2} < \frac{\varepsilon}{2}$. On the assumption $f(a_0) < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < f(a_m)$, we can find j minimal such that

$$f(a_j) < \frac{\varepsilon}{2} \quad \text{and} \quad -\frac{\varepsilon}{2} < f(a_{j+1}).$$

Finally, compare $f(a_j)$ with $-\varepsilon < -\frac{\varepsilon}{2}$ and $f(a_{j+1})$ with $\frac{\varepsilon}{2} < \varepsilon$. If $-\varepsilon < f(a_j)$, we have $|f(a_j)| < \varepsilon$. If $f(a_{j+1}) < \varepsilon$, we have $|f(a_{j+1})| < \varepsilon$. If both $f(a_j) < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < f(a_{j+1})$, then we would have $|f(a_{j+1}) - f(a_j)| > \varepsilon$, which contradicts $|a_{j+1} - a_j| \leq \delta$. □

Unlike the proofs of this result we know from the literature, the foregoing proof is entirely choice-free.[10] It seems to be characteristic of constructions of (uniquely determined or) approximate solutions that they do not even require countable choice [19].

[10] Again, this proof was taken from lecture notes of the second author.

3. The Fundamental Theorem of Algebra

In the framework of constructive analysis sketched above we now give a detailed proof of the fundamental theorem of algebra, essentially following Martin Kneser [9].[11] The point is that we present the proof in a choice-free form; this of course includes all auxiliary lemmata. Our exposition is rather detailed, because—as mentioned in the introduction—we want to make the algorithmic content of this proof as clear as possible, keeping in mind that it may later serve as a basis for extracting a program.

A somewhat different presentation of Kneser's proof has been given in [22]. However, it is stressed in ([22]: 437) that this proof "essentially rests on the axiom of choice".[12]

As usual, real numbers give rise to complex numbers, which in the present context are represented by pairs of modulated Cauchy sequences of rationals, and thus are also sequentially complete (Theorem 4).

Fix the degree $n \geq 1$ of the polynomials to be considered. The bulk of the technical work is in the proof of the following lemma.

Lemma 11. *We can find $q \in \mathbb{Q}$ with $0 < q < 1$ (depending only on our fixed $n \in \mathbb{N}$) such that for every monic polynomial $f(\xi) = \xi^n + a_{n-1}\xi^{n-1} + \cdots + a_1\xi + a_0 \in \mathbb{C}[\xi]$ of degree n and with $|a_0| > 0$ we can find $z \in \mathbb{C}$ satisfying*

$$|z| \leq \left(\frac{|a_0|}{q}\right)^{\frac{1}{n}} \quad \text{and} \quad |f(z)| \leq q|a_0|.$$

From the proof it will be clear that the approximate root $z \in \mathbb{C}$ we construct depends on the *representation* of the coefficients in $\mathbb{C}$; that is, on the choice of the pairs of modulated Cauchy sequences of rationals which represent them. This is to be expected; for instance, there is no continuous function $f : \mathbb{C} \to \mathbb{C}$ such that $f(z)^2 = z$; to resolve the non-uniqueness of the roots one needs Riemann surfaces.

Theorem 12 (Fundamental Theorem of Algebra). *For every monic polynomial $g(\xi) = \xi^n + b_{n-1}\xi^{n-1} + \cdots + b_1\xi + b_0 \in \mathbb{C}[\xi]$ of degree $n \geq 1$ we can find $w \in \mathbb{C}$ such that $g(w) = 0$.*

Proof. Choose $c > 0$ with $|b_0| \leq c$. Let q (depending on n) be supplied by Lemma 11 (for g in place of f). It clearly suffices to inductively construct a sequence $(w_m)_{m\in\mathbb{N}}$ of complex numbers such that

$$|w_{m+1} - w_m| \leq (q^{m-1}c)^{\frac{1}{n}} \tag{7}$$

[11] Once more, our version of this proof stems from the lecture notes of the second author. It has even been a basis of an implementation of the fundamental theorem of algebra in the interactive prover Coq [12], done by Barendregt, Geuvers, Pollack, Wiedijk and Zwanenburg [7].

[12] We also use the occasion to mention in print a small oversight in the proof in [22]—the specification of α on page 434 contains a circularity. This was first noted by Barendregt and Geuvers, and a correction can be found on Troelstra's home page.

and

$$|g(w_m)| \leq q^m c\,. \tag{8}$$

Any such sequence is a modulated Cauchy sequence in $\mathbb{C}$ whose limit is a root of g.

To start with, set $w_0 = 0$. For the induction step, assume that w_m has already been found; then $f_m(\xi) = g(w_m + \xi) \in \mathbb{C}[\xi]$ is a monic polynomial of degree n with constant coefficient $g(w_m)$. By comparing $|g(w_m)|$ with $0 < q^{m+1}c$, we either have $|g(w_m)| < q^{m+1}c$ or $|g(w_m)| > 0$ (Lemma 1). In the former case, we may take $w_{m+1} = w_m$. In the latter case, applying Lemma 11 to f_m yields $z \in \mathbb{C}$ such that

$$|z| \leq \left(\frac{|g(w_m)|}{q}\right)^{\frac{1}{n}} \leq (q^{m-1}c)^{\frac{1}{n}} \quad \text{and} \quad |f_m(z)| \leq q|g(w_m)| \leq q^{m+1}c.$$

(The second inequality of each pair holds according to (8).) So in this case we may set $w_{m+1} = w_m + z$, for which $g(w_{m+1}) = f_m(z)$. □

In this proof of Theorem 12, the tacit invocation of sequential completeness is fairly inessential, since—just as the z supplied by Lemma 11—each w_m can easily be constructed so that $w_m = a_m + ib_m$ with *rational* real and imaginary part a_m and b_m. (This is possible in virtue of the approximate character of each result.) As then (a_m) and (b_m) both are a modulated Cauchy sequences of rationals, they respectively represent real numbers a and b, so that $w = a + ib$ is the requested complex number with $g(w) = 0$.

In the induction step the cases of the crucial distinction may overlap, and thus a choice—which may even depend on the foregoing one—seems to be necessary whenever both cases happen simultaneously. However, Lemma 1 automatically makes this choice in every given situation. In other words, countable choice, let alone dependent choice, is not involved in the present proof of the fundamental theorem of algebra. Also, the proof of the fundamental theorem of algebra given by Ruitenburg in [16] is without any choice, and designed for complex numbers as represented by pairs of modulated Cauchy sequences of rationals. It is done by first proving the discrete fundamental theorem of algebra (the one for algebraic rather than complex numbers) by field-theoretic means (splitting fields, minimal polynomials, etc.), and then approximating complex numbers by algebraic ones.

Similar in strategy, Richman's choice-free approach [14] to an approximate version of the fundamental theorem of algebra appears to be independent of any specific presentation of the reals. Part of the price one has to pay for this generality is that one gets a bit less: a multiset of algebraic numbers as close as one pleases to the imagined multiset of complex roots. To extend this result to an exact version, by picking out elements from those multisets, only requires a fairly weak and classically valid fragment of countable choice [6].[13]

[13] Roughly speaking, this is countable choice with the additional presupposition that all choices but at most one—of which of course one does not know when it occurs—are choices (!) from a singleton set. It

Unlike these two proofs, the following one is truly elementary, which might be of importance to the form of an extracted program.

Proof of Lemma 11. Before going into the details, let us briefly sketch the approach that will prove successful in the following. First, we pick an appropriate one of the monomials $a_1\xi, \ldots, a_n\xi^n = \xi^n$, say $a_k\xi^k$, and next determine z such that a_0 and $a_k z^k$ differ by a negative factor. For $|z| = r$ we assume only that this monomial is sufficiently small, i.e., $|a_k|r^k \le |a_0| + \varepsilon$. Then we obtain (see (15))

$$|f(z)| \le |a_0 + a_k z^k| + \sum_{i\neq 0,k} |a_i z^i| \le |a_0| - |a_k|r^k + 2\varepsilon + \sum_{i\neq 0,k} |a_i| r^i.$$

So the point is to choose k and r in such a way that the monomials $|a_i|r^i$ $(i \neq 0, k)$ are small compared to $|a_k|r^k$, but on the other hand are not too small compared to $|a_0|$ (see (14)).

Step 0. To start with the construction, set

$$q = 1 - \frac{1}{4 \cdot 3^{2n^2-1}} \quad \text{and} \quad \varepsilon = \min\left\{ \frac{|a_0|}{2(n+1)} \cdot \frac{1}{4 \cdot 3^{2n^2-1}}, (q^{-1} - 1)|a_0| \right\}.$$

These choices will be explained in steps 6, 10 und 11 below.

Step 1. Let $m(s) := \max\{\, |a_i|s^i \mid i = 1, \ldots, n \,\}$ for every real s. Pick $t \in \mathbb{Q}$, $t > 0$, such that

$$|m(t) - |a_0|| \le \varepsilon. \tag{9}$$

This is possible by the approximate intermediate value theorem (Theorem 10), for $m(0) = 0 < |a_0|$ and, because $s^n \le m(s)$, there is an $s_0 > 0$ such that $|a_0| < m(s_0)$; note in this context that m is uniformly continuous over the compact interval $[0, s_0]$.

Step 2. For $i = 1, \ldots, n$ choose $y_i \in \mathbb{Q}$, $y_i > 0$, such that $|y_i - |a_i|t^i| \le \varepsilon$. With $y_{ij} := \frac{y_i}{3^{ij}}$ (for $j \in \mathbb{Z}$) we obtain

$$\left| y_{ij} - |a_i|\Big(\frac{t}{3^j}\Big)^i \right| = \left| \frac{y_i}{3^{ij}} - \frac{|a_i|t^i}{3^{ij}} \right| \le \frac{\varepsilon}{3^{ij}}. \tag{10}$$

Step 3. For every $j \in \mathbb{Z}$, pick $k_j \in \{1, \ldots, n\}$ minimal such that $y_{ij} \le y_{k_j j}$ for $i = 1, \ldots, n$. So

$$|a_i|\Big(\frac{t}{3^j}\Big)^i \le y_{ij} + \frac{\varepsilon}{3^{ij}} \le y_{k_j j} + \frac{\varepsilon}{3^{ij}} \le |a_{k_j}|\Big(\frac{t}{3^j}\Big)^{k_j} + \frac{\varepsilon}{3^{k_j j}} + \frac{\varepsilon}{3^{ij}}. \tag{11}$$

Step 4. We have

$$k_j \ge k_{j+1}. \tag{12}$$

is further noteworthy that many appeals to countable choice can be reduced to that weaker principle. We refer to [15] for more details.

Indeed, for $i \geq k_j$,

$$y_{k_j(j+1)} = \frac{1}{3^{k_j(j+1)}} y_{k_j} = \frac{1}{3^{k_j}} y_{k_j j} \geq_{(*)} \frac{1}{3^i} y_{ij} = \frac{1}{3^{i(j+1)}} y_i = y_{i(j+1)} ;$$

whence $k_{j+1} \leq k_j$ according to the choice of k_{j+1} ("least possible").

Step 5. Choose $j > 0$ minimal such that $k_{j-1} = k_j = k_{j+1} =: k$. Because of (12) we have

$$n \geq k_{-1} \geq k_0 \geq k_1 \geq \cdots \geq k_{j-1} = k_j = k_{j+1},$$

where out of each pair of two consecutive $\geq$ (except for the first one) at least one must be $>$. Hence for $2i - 1 \leq j$ we have $k_{2i-1} \leq n - i$.

Case 1. j is of the form $2i - 1$. Then

$$k_j \leq n - i = n - \frac{j+1}{2} \implies 2k_j \leq 2n - j - 1 \implies j \leq 2(n - k_j) - 1 .$$

Case 2. j is of the form $2i$. Then

$$k_j = k_{j-1} \leq n - i = n - \frac{j}{2} \implies 2k_j \leq 2n - j \implies j \leq 2(n - k_j) .$$

Hence generally we have

$$j \leq 2(n - k). \tag{13}$$

Step 6. Let $r := \frac{t}{3^j} \in \mathbb{Q}$. We estimate $|a_k| r^k$ against $|a_0|$. For $i = 1, \ldots, n$, by (11) and $k = k_j$ we have $|a_k| r^k = |a_k| (\frac{t}{3^j})^k \geq |a_i| (\frac{t}{3^j})^i - 2\varepsilon \geq \frac{1}{3^{nj}} |a_i| t^i - 2\varepsilon$, hence

$$\begin{aligned} |a_k| r^k &\geq \frac{1}{3^{nj}} m(t) - 2\varepsilon \quad \text{for } m(t) = \max\{\, |a_i| t^i \mid i = 1, \ldots, n \,\} \text{ as before} \\ &\geq \frac{1}{3^{nj}} |a_0| - 3\varepsilon \quad \text{by the choice of } t \text{ in (9).} \end{aligned}$$

By (13) we have $nj \leq 2n(n - k) \leq 2n^2 - k$ and therefore

$$|a_k| r^k \geq \frac{1}{3^{2n^2-k}} |a_0| - 3\varepsilon. \tag{14}$$

In particular $|a_k| > 0$ because of $\varepsilon < \frac{|a_0|}{3} \cdot \frac{1}{3^{2n^2-1}}$.

Step 7. We first determine $z \in \mathbb{C}$ with $|z| = r$ such that a_0 and $a_k z^k$ differ by a negative factor. Write $a_0 = |a_0| e^{i x_0}$ and $a_k = |a_k| e^{i x_k}$, and let $z = |r| e^{iy}$ with y such that $x_0 + \pi = x_k + ky$.

We show that $|a_0 + a_k z^k| \leq |a_0| - |a_k z^k| + 2\varepsilon$. Now $a_k z^k = -c a_0$ with $c \in \mathbb{R}$, $c > 0$, and by the choice of t in (9) also $|a_k z^k| \leq |a_0| + \varepsilon$. Therefore

$$c|a_0| = |a_k z^k| \leq |a_0| + \varepsilon \quad \text{and} \quad 0 < c \leq 1 + \frac{\varepsilon}{|a_0|},$$

hence

$$\begin{aligned}
|a_0 + a_k z^k| &= |a_0| \cdot |1 - c| \\
&= |a_0| \max(1 - c, c - 1) \\
&= |a_0| \max(1 - c, 1 - c + 2(c - 1)) \\
&\le |a_0| \max\Big(1 - c, 1 - c + 2\frac{\varepsilon}{|a_0|}\Big) \\
&= |a_0|(1 - c) + 2\varepsilon \\
&= |a_0| - |a_k z^k| + 2\varepsilon.
\end{aligned}$$

We now obtain

$$\begin{aligned}
|f(z)| &\le |a_0 + a_k z^k| + \sum_{i \neq 0,k} |a_i z^i| \\
&\le |a_0| - |a_k z^k| + 2\varepsilon + \sum_{i \neq 0,k} |a_i z^i| \\
&= |a_0| - |a_k| r^k + 2\varepsilon + \sum_{i \neq 0,k} |a_i| r^i.
\end{aligned} \tag{15}$$

Step 8. Estimate of $\sum_{1 \le i < k} |a_i| r^i$. Let $1 \le i < k$. Then

$$\begin{aligned}
|a_i| r^i &= 3^i |a_i| \Big(\frac{r}{3}\Big)^i \\
&= 3^i |a_i| \Big(\frac{t}{3^{j+1}}\Big)^i \\
&\le 3^i \Big(|a_k| \big(\frac{t}{3^{j+1}}\big)^k + \frac{\varepsilon}{3^{i(j+1)}} + \frac{\varepsilon}{3^{k(j+1)}}\Big) \quad \text{by (11), for } k = k_{j+1} \\
&\le 3^i \Big(|a_k| \big(\frac{t}{3^j}\big)^k \frac{1}{3^k} + 2\frac{\varepsilon}{3^i}\Big) \\
&= \frac{1}{3^{k-i}} |a_k| r^k + 2\varepsilon,
\end{aligned}$$

hence (using $\frac{1}{3^{k-1}} + \cdots + \frac{1}{3} = \frac{1}{2}\big(1 - \frac{1}{3^{k-1}}\big)$)

$$\sum_{1 \le i < k} |a_i| r^i \le |a_k| r^k \frac{1}{2}\Big(1 - \frac{1}{3^{k-1}}\Big) + 2(k-1)\varepsilon. \tag{16}$$

Step 9. Estimate of $\sum_{k < i \le n} |a_i| r^i$. Let $k < i \le n$. Then

$$\begin{aligned}
|a_i| r^i &= \frac{1}{3^i} |a_i| (3r)^i \\
&= \frac{1}{3^i} |a_i| \Big(\frac{t}{3^{j-1}}\Big)^i \\
&\le \frac{1}{3^i} \Big(|a_k| \big(\frac{t}{3^{j-1}}\big)^k + \frac{\varepsilon}{3^{i(j-1)}} + \frac{\varepsilon}{3^{k(j-1)}}\Big) \quad \text{by (11), for } k = k_{j-1}
\end{aligned}$$

$$\leq \frac{1}{3^i}\Big(|a_k|\Big(\frac{t}{3^j}\Big)^k \cdot 3^k + 2\cdot 3^k\varepsilon\Big)$$
$$\leq \frac{1}{3^{i-k}}|a_k|r^k + 2\varepsilon,$$

hence (using $\frac{1}{3}+\cdots+\frac{1}{3^{n-k}} = \frac{1}{2}\big(1-\frac{1}{3^{n-k}}\big)$)

$$\sum_{k<i\leq n}|a_i|r^i \leq |a_k|r^k\frac{1}{2}\Big(1-\frac{1}{3^{n-k}}\Big) + 2(n-k)\varepsilon. \tag{17}$$

Step 10. From (15), (16) and (17) one obtains

$$|f(z)| \leq |a_0| - |a_k|r^k + 2\varepsilon + \sum_{i\neq 0,k}|a_i|r^i$$
$$\leq 2n\varepsilon + |a_0| - |a_k|r^k + |a_k|r^k\frac{1}{2}\Big(1-\frac{1}{3^{k-1}}\Big) + |a_k|r^k\frac{1}{2}\Big(1-\frac{1}{3^{n-k}}\Big)$$
$$\leq 2n\varepsilon + |a_0| - \frac{1}{2}\cdot\frac{1}{3^{k-1}}|a_k|r^k$$
$$\leq 2n\varepsilon + |a_0| - \frac{1}{2\cdot 3^{k-1}}\Big(\frac{1}{3^{2n^2-k}}|a_0| - 3\varepsilon\Big) \quad \text{by (14)}$$
$$\leq 2(n+1)\varepsilon + \Big(1-\frac{1}{2\cdot 3^{2n^2-1}}\Big)|a_0|$$
$$\leq \underbrace{\Big(1-\frac{1}{4\cdot 3^{2n^2-1}}\Big)}_{=q}|a_0| \quad \text{for } \varepsilon \leq \tfrac{|a_0|}{2(n+1)}\cdot\tfrac{1}{4\cdot 3^{2n^2-1}}$$

Step 11. Observe $|z| = r$, hence

$$|z|^n = r^n \leq t^n \leq m(t) \leq |a_0| + \varepsilon \leq q^{-1}|a_0|$$

because of $\varepsilon \leq (q^{-1}-1)|a_0|$. Therefore we have $|z| \leq (q^{-1}|a_0|)^{\frac{1}{n}}$, as was required. □

4. Dedekind Reals and Order Completeness

By a *Dedekind real* we understand a located Dedekind cut in the rationals: that is, a disjoint pair (L, U) of inhabited open subsets of $\mathbb{Q}$ that is *cut located* inasmuch as either $p \in L$ or $q \in U$ for all $p, q \in \mathbb{Q}$ with $p < q$. In particular, $p < q$ whenever $p \in L$ and $q \in U$. The *lower cut* L and the *upper cut* U uniquely determine each other, as one is the open interior of the complement of the other.[14] Needless to say, every rational r is identified with the Dedekind real $(\{\, p \in \mathbb{Q} \mid p < r \,\}, \{\, q \in \mathbb{Q} \mid r < q \,\})$.

[14] Although it may seem redundant, we have chosen to define a Dedekind real as a pair rather than just as either a lower or an upper cut. Among our reasons for doing so is that pairs are better suited when it comes to considering *generalized* real numbers, i.e., not necessarily located Dedekind cuts, as, e.g., in [13, 17].

Every lower cut L (or upper cut U) is bounded from above (or from below) and *downwards monotone* (or *upwards monotone)*: that is, for all $p, q \in \mathbb{Q}$ with $p < q$, if $q \in L$, then $p \in L$ (or, if $p \in U$, then $q \in U$). Moreover, every Dedekind real is *approximable*: that is, for every rational $\varepsilon > 0$ there are $p \in L$ and $q \in U$ with $q - p < \varepsilon$. For the latter observation, and for the second part of the following, we may refer to ([22]: Chapter 5, Section 5); the first part is routine to verify.

Lemma 13.

(a) *If L is an open subset of $\mathbb{Q}$ and U denotes the open interior of $\mathbb{Q} \setminus L$, then (L, U) is cut located if and only if L is downwards monotone and order located from above.*

(b) *The following items are equivalent for a pair (L, U) of subsets of $\mathbb{Q}$:*

  (i) *L is inhabited, U is inhabited, and (L, U) is cut located.*

  (ii) *L is downwards monotone, U is upwards monotone, and (L, U) is approximable.*

The strict partial order of Dedekind reals is given by $(L, U) < (L', U')$ if and only if $L' \cap U$ is inhabited. So $\leq$, $\neq$, and $=$ could be introduced—for Dedekind reals just as for any other type of reals—as the negation of $<$, the disjunction of $<$ and $>$, and the conjunction of $\geq$ and $\leq$, respectively. For Dedekind reals, however, all these relations admit of a simpler definition: namely, $(L, U) \leq (L', U')$ as $L \subset L'$ or, equivalently, as $U \supset U'$; $(L, U) = (L', U')$ as $L = L'$ or, equivalently, as $U = U'$; $(L, U) \neq (L', U')$ as $L' \cap U$ or $L \cap U'$ being inhabited. Referring to ([22]: Chapter 5, Section 5) for further details, so for arithmetic, we write $\mathbb{D}$ for the set of Dedekind reals.

Let $\mathfrak{R}$ be an Archimedean ordered and valued Heyting field: that is, a model of all the axioms—for the moment, except for order completeness—that were listed by Bridges in [4].[15] (In compliance with Theorem 5 above, we say that any such $\mathfrak{R}$ is *order complete* whenever the statement of that theorem is valid for this $\mathfrak{R}$ in place of $\mathbb{R}$.) These axioms—together with order completeness, of course—embody all the properties of real numbers, such as the approximate splitting principle (cf. Lemma 1), that are commonly accepted in constructive mathematics. In particular, any such $\mathfrak{R}$ contains $\mathbb{Q}$ as an ordered subfield that is *dense* in the sense that for all $x, y \in \mathfrak{R}$ with $x < y$ there is $q \in \mathbb{Q}$ such that $x < q < y$.

Both the Dedekind reals and the real numbers as defined earlier in this paper, by way of modulated Cauchy sequences of rationals, form a perfect model of those axioms. As we have seen above and shall remember below, this includes order completeness.

[15] There and in other places order completeness is called the "(constructive) least-upper-bound principle". See also [18] for a case study of real numbers as black boxes that, among other things, serves to justify the choice of these axioms.

If $S \subset \mathfrak{R}$ is inhabited, bounded from above, and order located from above, then $\xi_S = (L_S, U_S)$ is a Dedekind real with L_S and U_S as the open interiors of

$$\{\, p \in \mathbb{Q} \mid p < r \text{ for some } r \in S \,\} \quad \text{and} \quad \{\, q \in \mathbb{Q} \mid s \leq q \text{ for all } s \in S \,\},$$

respectively. In particular, we can assign a Dedekind real $\pi(x)$ to each $x \in \mathfrak{R}$ by setting

$$\pi(x) = (\{\, p \in \mathbb{Q} \mid p < x \,\}, \{\, q \in \mathbb{Q} \mid x < q \,\}) \ .$$

Then $\pi(q) = q$ for all rationals q, and $\pi(x) < \pi(x')$ precisely when $x < x'$; whence we actually have defined an injective mapping $\pi : \mathfrak{R} \to \mathbb{D}$.

In other words, $\mathfrak{R}$ is canonically embedded into $\mathbb{D}$, which rephrases the saying that there are *at least* as many Dedekind reals as there are real numbers of an arbitrary kind like $\mathfrak{R}$. The question to which we now turn is under which circumstances there are *exactly* as many, or rather—in view of the canonical character of π—when π is a one-to-one correspondence.

In [19] it is proved,[16] without invoking any choice principle, that $\mathfrak{R}$ is order complete if and only if a mapping $\sigma : \mathbb{D} \to \mathfrak{R}$ exists that preserves $<$ and operates as the identity on $\mathbb{Q}$. In fact, if $\xi = (L, U) \in \mathbb{D}$, then $L \subset \mathbb{Q}$ —considered as a subset of $\mathfrak{R}$—is inhabited, bounded from above, and order located from above; whence the supremum of L exists provided that $\mathfrak{R}$ is order complete. Now it is routine to verify that

$$\sigma : \mathbb{D} \to \mathfrak{R}, \ \xi = (L, U) \mapsto \sup L$$

is a mapping with $\sigma(\xi) < \sigma(\xi')$ precisely when $\xi < \xi'$, and with $\sigma(q) = q$ for all $q \in \mathbb{Q}$. Conversely, if we are given such a mapping $\sigma : \mathbb{D} \to \mathfrak{R}$, and if $S \subset \mathfrak{R}$ is inhabited, bounded from above, and order located from above, then $\sigma(\xi_S)$—with ξ_S as defined before—is a supremum of S.

As a consequence, the Dedekind reals are order complete: the identity mapping on $\mathbb{D}$ clearly meets the specifications of such a σ. Moreover, if $\mathfrak{R}$ is order complete and thus $\sigma : \mathbb{D} \to \mathfrak{R}$ can be defined as above, it automatically is the two-sided inverse of π, so that if (and, of course, only if) $\mathfrak{R}$ is order complete, then $\mathfrak{R}$ is order isomorphic to the Dedekind reals with $\mathbb{Q}$ being invariant under the isomorphism. So $\mathbb{D}$ can legitimately be characterized as the order completion of $\mathbb{Q}$ inasmuch as the statement of Theorem 5 is the right way to phrase order completeness—which it definitely is from the constructive perspective.

Troelstra and van Dalen stated in ([22]: Chapter 5, 5.12.iii) that, roughly speaking, $\mathbb{D}$ is—up to order isomorphism—the one and only ordered extension of $\mathbb{Q}$ in which every lower cut in the rationals possesses a least upper bound. This characterization of $\mathbb{D}$ turns out to be equivalent to the one we gave above by way of order completeness. In fact, the former follows from the latter, because every lower cut satisfies the hypotheses of order completeness: according to the first part of Lemma 13, it is inhabited, bounded

[16] This goes back to an idea of Fred Richman (personal communication to the first author).

from above, and order located from above. For the reverse implication, it suffices to notice that if L_S is assigned to $S \subset \mathfrak{R}$ as before, and if $\sup L_S \in \mathfrak{R}$ exists, then it is also the supremum of S.

In retrospect, we have achieved the following characterization.

Theorem 14. *The following items all are equivalent:*

(a) $\mathfrak{R}$ *is order complete.*

(b) *Every lower cut in* $\mathbb{Q}$ *has a supremum in* $\mathfrak{R}$.

(c) *There is a mapping* $\sigma : \mathbb{D} \to \mathfrak{R}$ *that preserves* $<$ *and each rational.*

(d) *The canonical mapping* $\pi : \mathfrak{R} \to \mathbb{D}$ *is bijective.*

(e) $\mathfrak{R}$ *is order isomorphic to* $\mathbb{D}$.

In view of Theorem 5, we have the following

Corollary 15. $\mathbb{R}$ *is order isomorphic to* $\mathbb{D}$.

Note that this does not require countable choice; we have stressed the presence of witnesses throughout.

References

[1] Aczel, Peter and Michael Rathjen: 2002. *Notes on Constructive Set Theory*. Draft of an extended and revised version of the preprint available at the Mittag-Leffler Institute.

[2] Bishop, Errett: 1967. *Foundations of Constructive Analysis*. New York: McGraw-Hill.

[3] Bishop, Errett and Douglas Bridges: 1985. *Constructive Analysis*. Grundlehren der mathematischen Wissenschaften 279. Berlin: Springer.

[4] Bridges, Douglas: 1999. Constructive mathematics: a foundation for computable analysis. *Theoret. Comput. Sci.* 219: 95–109.

[5] Bridges, Douglas and Fred Richman: 1987. *Varieties of Constructive Mathematics*. Cambridge: Cambridge University Press.

[6] Bridges, Douglas, Fred Richman, and Peter Schuster: 2000. A weak countable choice principle. *Proc. Amer. Math. Soc.* 128: 2749–2752.

[7] Geuvers, Herman, Randy Pollack, Freek Wiedijk, and Jan Zwanenburg: June 2000. *Theory Development Leading to the Fundamental Theorem of Algebra*. Manuscript (automatically generated).

[8] Ishihara, Hajime and Peter Schuster: 2002. A constructive uniform continuity theorem. *Quart. J. Math.* 53: 185–193.

[9] Kneser, Martin: 1981. Ergänzung zu einer Arbeit von Hellmuth Kneser über den Fundamentalsatz der Algebra. *Math. Z.* 177: 285–287.

[10] Mines, Ray, Fred Richman, and Wim Ruitenburg: 1987. *A Course in Constructive Algebra*. Berlin: Springer.

[11] Palmgren, Erik: 2003. *Constructive Completions of Ordered Sets, Groups and Fields*. Preprint, Uppsala University.

[12] Paulin-Mohring, Christine and Benjamin Werner: 1993. Synthesis of ML programs in the system Coq. *J. Symbolic Computation* 11: 1–34.

[13] Richman, Fred: 1998. Generalized real numbers in constructive mathematics. *Indag. Mathem. (N.S.)* 9: 595–606.

[14] Richman, Fred: 2000. The fundamental theorem of algebra: a constructive development without choice. *Pacific J. Math.* 196: 213–230.

[15] Richman, Fred: 2001. Constructive mathematics without choice. In [20]: 199–205.

[16] Ruitenburg, Willem B. G.: 1991. Constructing roots of polynomials over the complex numbers. In: A. M. Cohen (ed.), *Computational Aspects of Lie Group Representations and Related Topics. Proc. 1990 Computer Algebra Seminar*, CWI Tract 84, 107–128, Amsterdam: Centrum voor Wiskunde en Informatica.

[17] Schuster, Peter: 2000. A constructive look at generalized Cauchy reals. *Math. Logic Quart.* 46: 125–134.

[18] Schuster, Peter: 2002. Real numbers as black boxes. *New Zealand J. Math.* 31: 189–202.

[19] Schuster, Peter: 2003. Unique existence, approximate solutions, and countable choice. *Theoret. Comput. Sci.* 305: 433-455.

[20] Schuster, Peter, Ulrich Berger, and Horst Osswald (eds.): 2001. *Reuniting the Antipodes—Constructive and Nonstandard Views of the Continuum*. Synthese Library vol. 306. Dordrecht: Kluwer.

[21] Taschner, Rudolf: 1991, 1992, 1993. *Lehrgang der konstruktiven Mathematik*. Drei Bände. Wien: Manz and Hölder-Pichler-Tempsky; 2nd ed. of vols. 1,2, ibid., 1995.

[22] Troelstra, Anne S. and Dirk van Dalen: 1988. *Constructivism in Mathematics. An Introduction*, vols. 121, 123. Studies in Logic and the Foundations of Mathematics. Amsterdam: North-Holland.

Mathematisches Institut
Universität München
Theresienstraße 39
80333 München
Germany

E-mail: pschust@mathematik.uni-muenchen.de
schwicht@mathematik.uni-muenchen.de

Russell's Paradox in Consistent Fragments of Frege's *Grundgesetze der Arithmetik*

Kai F. Wehmeier

Abstract. We provide an overview of consistent fragments of the theory of Frege's *Grundgesetze der Arithmetik* that arise by restricting the second-order comprehension schema. We discuss how such theories avoid inconsistency and show how the reasoning underlying Russell's paradox can be put to use in an investigation of these fragments.

1. Introduction.

On June 16th, 1902, Russell communicates 'his' paradox to Frege:

> Let w be the predicate of being a predicate that cannot be predicated of itself. Can one predicate w of itself? From any answer follows its opposite. Therefore one must infer that w is not a predicate. Likewise, there is no class (as a whole) of those classes that as wholes do not belong to themselves. From this I infer that under certain circumstances a definable set does not form a whole.[1]

In a postscript to this letter, Russell uses Peano's notation to give the first symbolic expression of the paradox:

$$w = \mathrm{cls} \cap x \ni (x \sim \epsilon\, x).\ \supset :\ w \epsilon\, w.\ =\ .w \sim \epsilon\, w.^{2}$$

One would imagine that Frege was not much impressed by the predicability version of the antinomy that Russell mentions first (which is basically the heterologicality paradox), as it is obviated, within the *Begriffsschrift*, by the stratification of functions into levels. The class version of the contradiction, however, nowadays known as

[1] Cf. [14]: 211 (the English translation is mine): 'Sei w das Prädicat, ein Prädicat zu sein welches von sich selbst nicht prädicirt werden kann. Kann man w von sich selbst prädiciren? Aus jeder Antwort folgt das Gegentheil. Deshalb muss man schliessen dass w kein Prädicat ist. Ebenso giebt es keine Klasse (als Ganzes) derjenigen Klassen die als Ganze sich selber nicht angehören. Daraus schliesse ich dass unter gewissen Umständen eine definierbare Menge kein Ganzes bildet.'

[2] See [14]: 212. In modern notation, Russell's formula reads as follows:
$w = \{x : \neg x \in x\} \to (w \in w \leftrightarrow \neg w \in w)$.

Russell's paradox, Frege immediately recognised as being reconstructible within the theory of *Grundgesetze*. He regarded the occurrence of this contradiction as a serious blow to his life's work—one can clearly sense his horror between the lines of his answer to Russell, written on June 22nd, 1902:

> Your discovery of the contradiction utterly surprised and, I am almost inclined to say, dismayed me, as the foundation upon which I believed arithmetic to be built, thereby begins to rock. It thus seems that the transformation of the generality of an identity into an identity of value-ranges (§9 of my Grundgesetze) is not always permitted, that my law V (§20 p. 36) is false and that my explanations in §31 do not suffice to secure my combinations of signs a denotation in all cases.[3]

It is clear from these remarks that Frege himself considered his *Grundgesetz* V as the source of the antinomy. This is a very reasonable diagnosis, because basic law V figures prominently in the derivation of the contradiction, as we shall see shortly. Most philosophers have followed Frege in this assessment. Michael Dummett has pointed out, however, that another contentious principle plays an equally prominent role in the genesis of the contradiction, viz., Frege's substitution rule for 'free' second-order variables ([12]: §48, subsection 9, 62–63). Such a rule is equivalent, under certain conditions, to the full (impredicative) second-order comprehension schema, and it is this impredicativity that Dummett takes to be responsible for the contradiction.[4] The bulk of this paper will be concerned with the problem of weakening the comprehension schema in such a way as to obtain a consistent fragment[5] of Frege's original theory.[6]

To set the stage, let us begin by considering the version of Russell's paradox that Frege discusses first in the *Nachwort* to the second volume of *Grundgesetze* ([13]: 256–257) (as reconstructed in terms of axiomatic second-order logic). Recall that, according to Frege, there is, associated with any second-order entity (concept) F, a certain first-order entity (object), the concept's extension (course-of values, value-

[3] Cf. [14]: 213 (the English translation is mine): 'Ihre Entdeckung des Widerspruchs hat mich auf's Höchste überrascht und, fast möchte ich sagen, bestürzt, weil dadurch der Grund, auf dem ich die Arithmetik sich aufzubauen dachte, in's Wanken geräth. Es scheint danach, dass die Umwandlung der Allgemeinheit einer Gleichheit in eine Werthverlaufsgleichheit (§9 meiner Grundgesetze) nicht immer erlaubt ist, dass mein Gesetz V (§20 S. 36) falsch ist und dass meine Ausführungen im §31 nicht genügen, in allen Fällen meinen Zeichenverbindungen eine Bedeutung zu sichern.'

[4] Cf. [9]: chapter 17 and, for critical discussion, [4].

[5] We shall, in fact, not work with the term logic of Frege's *original* theory, but rather with reconstructions of this theory in terms of second-order predicate logics. This deviation from Frege's setting would seem to be harmless as regards the source of the Russell antinomy.

[6] John Burgess, in his forthcoming monograph [7], discusses this approach (as well as others) to fixing Frege's logic in much greater detail than is here possible. Much interesting and important work has been done on consistent modifications of Frege's systems in *Grundlagen* and *Grundgesetze*, e.g., by G. Boolos, K. Fine, R. Hale, R. Heck, C. Wright, and E. Zalta; but insofar as their systems are not fragments *strictu sensu* of the *Grundgesetze* theory, their discussion falls outside the scope of the present note. The interested reader should consult, e.g., the collection of papers in [8], [11], and [1].

range) $(\hat{x})F(x)$. These extensions are governed by *Grundgesetz* V:

$$\forall F \forall G \left((\hat{x})F(x) = (\hat{y})G(y) \leftrightarrow \forall z(F(z) \leftrightarrow G(z)) \right).$$

That is, for any concepts F and G, their extensions $(\hat{x})F(x)$ and $(\hat{y})G(y)$ coincide if and only if the same objects fall under F as fall under G. Thus, basic law V is essentially an extensionality axiom for value-ranges.

Russell's paradox now arises as follows: Consider the concept R such that an object a falls under R if and only if a is the value-range of some concept F under which a does not fall. By (impredicative) second-order comprehension, such a concept R exists:

$$\exists R \, \forall a \left(R(a) \leftrightarrow \exists F(a = (\hat{z})F(z) \wedge \neg F(a)) \right).$$

Now let r be the extension of R, $r = (\hat{y})R(y)$. Suppose that r falls under R. By definition of R, r is then the extension of some F such that r does not fall under F, i.e., $\exists F(r = (\hat{z})F(z) \wedge \neg F(r))$. So $(\hat{z})F(z) = r = (\hat{y})R(y)$. Using basic law V, we find that $\forall x(F(x) \leftrightarrow R(x))$, that is, F and R are coextensional. Hence, since r does not fall under F, we also have $\neg R(r)$. Cancelling the assumption that r falls under R, we have shown that $R(r) \rightarrow \neg R(r)$, hence $\neg R(r)$. By definition of R, this means that for any concept F, if r is the extension of F, then r falls under F, i.e., $\forall F(r = (\hat{z})F(z) \rightarrow F(r))$. By specialising the universally quantified second-order variable F to R (R is in the second-order domain of quantification by the above instance of the comprehension schema), and noting that r is the extension of R, we obtain $R(r)$. So $\neg R(r)$ and $R(r)$—*voilà* the contradiction.

Let us now introduce a symbol for the membership relation by the following definition:

$$x \in y \leftrightarrow \forall F(y = (\hat{z})F(z) \rightarrow F(x)),$$

that is, x is a member of y if and only if, whenever y is the extension of some concept F, x falls under F. We can then immediately obtain the better known class version of Russell's paradox from the derivation above: Russell's condition $x \notin x$ transforms, upon elimination of the defined notion, into $\exists F(x = (\hat{z})F(z) \wedge \neg F(x))$, that is, into x's falling under the concept R introduced above, and the Russell class $\{y : y \notin y\}$ is just r, the extension of R.[7]

As is well known, Russell's paradox is closely related to Cantor's theorem. In fact, Russell himself discovered his paradox through an analysis of Cantor's theorem.[8] Cantor's theorem says that there is no bijection between the first- and the second-order entities (or, in more familiar terms, between the members of a given set A and

[7] Frege himself discusses (a variant of) this version of the paradox in the *Nachwort* ([13]: 257), making use of his analogue of the membership relation introduced in §34 (which is in fact, due to the peculiar nature of Frege's system, an application function). We speak of a 'variant' here because Frege defines the membership relation $x \in y$ as $\exists F(y = (\hat{z})F(z) \wedge F(x))$ (this definition is slightly more natural than the one used above: if y is not a value-range, then, according to the latter definition, it has no members, whereas according to the one we have given, everything will then be a member of y).

[8] Cf. Russell's letter to Frege of June 24th, 1902 ([14]: 215-6).

the subsets of A). The usual proof of this result shows, more specifically, that there is no map from the first- *onto* the second-order universe. Alternatively, in order to prove Cantor's theorem, one can show that there is no *1-1* map from the second- into the first-order universe. It is this alternative proof that reveals the connection to Russell's paradox: Suppose that $G \mapsto G^*$ is a 1-1 function from the second- into the first-order domain. Now consider the concept C with the following comprehension property:

$$C(x) \leftrightarrow \exists F(x = F^* \wedge \neg F(x)).$$

C^* is a first-order entity, so we may ask whether C^* falls under C. Suppose it does. Then by definition of C, there is a concept F such that $C^* = F^*$ and $\neg F(C^*)$. As $G \mapsto G^*$ is 1-1, F and C are the same, and so $\neg C(C^*)$. Cancelling the supposition that $C(C^*)$, we obtain $C(C^*) \rightarrow \neg C(C^*)$, i.e., $\neg C(C^*)$. Again by definition of C, this means that $\forall F(C^* = F^* \rightarrow F(C^*))$, from which $C(C^*)$ immediately follows. As $C(C^*)$ and $\neg C(C^*)$ cannot both be true, we must conclude that $G \mapsto G^*$ is not 1-1.

It should be clear that, in this version, Cantor's theorem basically *is* Russell's paradox: Reading $(\hat{x})G(x)$ instead of G^* and noting that basic law V claims this function to be 1-1, the proofs are literally identical.[9] It is the more surprising that Frege failed to realize at the outset that his value-range function would lead to inconsistency, as he was clearly aware of Cantor's theorem: In §164 of the second volume of *Grundgesetze*, he mentions that there are more classes of natural numbers than natural numbers, without crediting this insight to Cantor:

> Now an infinite number, which we have called Endless, belongs to the concept *finite number*; but this infinity does not yet suffice. If we call the extension of a concept that is subordinate to the concept *finite number* a *class of finite numbers*, then an infinite number that is greater than Endless belongs to the concept *class of finite numbers*; i.e., the concept *finite number* can be mapped into the concept *class of finite numbers*, but not vice versa the latter into the former.[10]

It remains puzzling why Frege did not see the consequences of Cantor's result for his own logical system. In any case, Russell did, and we shall now discuss how to avoid the antinomy by weakening the second-order comprehension schema.

[9] Cf. [5] for an illuminating discussion of these matters.

[10] Cf. [13]: 161 (the translation is mine): 'Nun kommt ja dem Begriffe *endliche Anzahl* eine unendliche Anzahl zu, die wir Endlos genannt haben; aber diese Unendlichkeit genügt noch nicht. Nennen wir den Umfang eines Begriffes, der dem Begriffe *endliche Anzahl* untergeordnet ist, eine *Klasse endlicher Anzahlen*, so kommt dem Begriffe *Klasse endlicher Anzahlen* eine unendliche Anzahl zu, die grösser als Endlos ist; d.h. es lässt sich der Begriff *endliche Anzahl* abbilden in den Begriff *Klasse endlicher Anzahlen*, aber nicht umgekehrt dieser in jenen.' It seems worth noting that Frege here deviates from the terminology he introduced in volume I: We can clearly map the sets of natural numbers into the natural numbers—by mapping every set to 0, say. But there is no 1-1 map from the sets of numbers into the numbers, so Frege presumably means 'map 1-1 into' when he says 'map into'. Or perhaps the vice versa clause is intended to mean that, whenever a relation maps the natural numbers into the sets of natural numbers, the converse of that relation will not map the sets into the numbers. The point seems marginal, but Frege's sloppiness is somewhat surprising.

2. Restricting the Comprehension Schema

We saw in the informal derivation of Russell's antinomy above that a certain instance of the comprehension schema plays a pivotal role in the proof, viz.,

$$\exists R\, \forall x \left(R(x) \leftrightarrow \exists F(x = (\hat{z})F(z) \wedge \neg F(x))\right).$$

It seems natural to ask what happens when comprehension is not available, or available only for formulae of a complexity below that of $\exists F(x = (\hat{z})F(z) \wedge \neg F(x))$. Peter Schroeder-Heister [18] conjectured that, in the complete absence of comprehension, that is, in the first-order fragment of Frege's theory, the antinomy would no longer be derivable. This was later confirmed by Terence Parsons [17]. We shall consider this system in subsection 2.1. Parsons' consistency proof was then extended, by Richard Heck [16], to the predicative fragment of Frege's theory, where the comprehension formulae must not contain any second-order quantifiers. We shall discuss Heck's theory, and some observations concerning it made in [19], in subsection 2.2. Subsequently, we briefly turn to a discussion of the Δ^1_1-CA fragment of Frege's theory that was recently proven consistent by Fernando Ferreira and the present author [10]. In the final subsection 2.4, we consider a related, though weaker theory proved consistent and discussed in [19], and contrast its linguistic setup with that of the theories discussed earlier.

2.1. The First-Order Fragment

We shall follow Heck [16] rather than Parsons [17] in setting out the first-order fragment. This is for reasons of convenience, as Parsons conscientiously works with a Fregean term logic, whereas the systems to be considered later are all based on predicate logic. In a first-order theory, we can obviously not formulate basic law V with the help of second-order universal quantification, as we did above. Thus we consider here schema V instead, that is, all instances of the schema

$$(\hat{x})\phi(x) = (\hat{y})\psi(y) \leftrightarrow \forall z(\phi(z) \leftrightarrow \psi(z)),$$

where ϕ and ψ are any formulae of the first-order fragment's language L_1. The terms and formulae of L_1 are generated by the following inductive definition:

1. Every individual variable is a term.
2. If s and t are terms, then $s = t$ is a formula.
3. Boolean combinations of formulae are formulae.
4. If x is an individual variable and $\phi(x)$ is a formula, then $\forall x\phi(x)$ is a formula and $(\hat{x})\phi(x)$ is a term (a value-range or VR term, as we shall say).

Parsons builds a structure for this language satisfying all instances of schema V as follows. Take ω as the domain of quantification. Assign the closed VR terms (where parameters from ω are allowed) natural numbers as ranks according to the rule: If $\phi(x)$ contains no VR terms, then the rank of $(\hat{x})\phi(x)$ is 0; if it does contain VR terms, and the maximal rank of a VR term occurring in it is n, then the rank of $(\hat{x})\phi(x)$ is $n+1$. Within a given rank, order the closed VR terms arbitrarily into an ω-sequence. Partition ω (or some infinite subset of it) into infinitely many infinite *candidate sets* U_i ($i < \omega$). Now assign values from $\bigcup_{i=0}^{\infty} U_i$ to the closed VR terms by recursion on ω^2. Suppose you are treating the mth VR term of rank n, $(\hat{x})\phi(x)$, say. All closed instances of VR terms occurring in $\phi(x)$ have rank lower than n and have thus already been assigned values. It is therefore determined whether for some VR term $(\hat{x})\psi(x)$ treated at some earlier stage we have $\forall x(\phi(x) \leftrightarrow \psi(x))$. If this is the case, we assign to $(\hat{x})\phi(x)$ whatever was assigned to $(\hat{x})\psi(x)$ before. Otherwise, we assign it the first element of U_n not yet used. In this way, we clearly obtain a model of the full schema V.[11]

We note in passing that John Burgess [6] has recently given a more constructive proof of Parsons' result. Furthermore, Warren Goldfarb [15] has shown that the first-order fragment is recursively undecidable, so it is not completely trivial mathematically (although it is unknown whether the theory is *essentially* undecidable). However, it does not seem likely that any interesting mathematics can be developed within this fragment.

2.2. The Predicative Fragment

The language L_2 of Heck's predicative fragment **H** results from Parsons' L_1 by adding second-order variables and quantifiers. That is, the terms and formulae of L_2 are generated by the clauses given above for L_1, supplemented by a clause for building atomic formulae $F(t)$ from second-order variables F and terms t, and closure under universal second-order quantification. Schema V is as before, where of course the instances are now formed by inserting arbitrary L_2-formulae for ϕ and ψ. In addition to schema V, **H** has as axioms all instances of predicative comprehension, that is, all instances of

$$\exists F \forall x(F(x) \leftrightarrow \phi(x)),$$

where ϕ contains no second-order quantifier. The deductive apparatus of **H** may be taken to be that of two-sorted first-order logic, where the objects and concepts represent the sorts.

[11] An anonymous referee asked whether the first-order fragment has finite models. This is not the case: Any model must contain values for $(\hat{x})(x = x)$ and $(\hat{x})(x \neq x)$; since no model is empty, by schema V, these two VR terms must denote distinct objects in any model. But then, $(\hat{y})(y = (\hat{x})(x = x))$, $(\hat{y})(y = (\hat{x})(x \neq x))$, and $(\hat{x})(x = x)$ must all have distinct values by schema V, etc.

We start by building a model for the first-order fragment of **H** by Parsons' procedure, assigning values to closed VR terms containing no second-order variables (but possibly parameters from ω), where we take care to choose the U_i in such a way that $\bigcup_{i=0}^{\infty} U_i$ has an infinite complement. We expand the resulting first-order structure to a second-order structure by letting the second-order quantifier range over the first-order definable subsets of ω. Now we assign values to those closed VR terms of L_2 that contain second-order parameters, but no second-order quantifiers, simply by expanding the parameters to their first-order definitions and choosing the value that was assigned to the corresponding first-order VR term. At this stage, we have stipulated enough in order for the predicative comprehension schema to be valid, as is easy to see. It remains to take care of those VR terms that contain, besides possibly first- and second-order parameters, second-order quantifiers. This can be done more or less as in the original Parsons procedure, using the infinite complement of $\bigcup_{i=0}^{\infty} U_i$ as values for the properly impredicative VR terms. It is then easy to check that all instances of schema V hold.

As was observed in [19], the theory **H** has a curious property: it proves the existence of objects that are not value-ranges. That is, the sentence $\exists x\, \forall G \left(x \neq (\hat{y})G(y)\right)$ is a theorem of **H**. There is even a term of L_2 witnessing this existential sentence, viz., the term r for the Russell class introduced above: $(\hat{z})(\exists F(z = (\hat{v})F(v) \wedge \neg F(z)))$. In other words, there is a value-range term of L_2 of which **H** proves that it does not *denote* a value-range. The proof is simple, and makes no use of the comprehension schema at all:

Argue in **H**. As before, let $R(x)$ be the formula $\exists F(x = (\hat{v})F(v) \wedge \neg F(x))$. Now assume $R(r)$, that is, $\exists F(r = (\hat{v})F(v) \wedge \neg F(r))$. Take such an F. Since $(\hat{v})F(v) = r = (\hat{z})R(z)$, by the appropriate instance of schema V, $\forall x(F(x) \leftrightarrow R(x))$. So, since $\neg F(r)$, we also have $\neg R(r)$. As above, we cancel the assumption and conclude $R(r) \rightarrow \neg R(r)$, i.e., $\neg R(r)$. This means $\forall F(r = (\hat{v})F(v) \rightarrow F(r))$. Now suppose that for some second-order G, $r = (\hat{y})G(y)$. It follows that $G(r)$ holds; but by schema V, it also follows that $\forall x(R(x) \leftrightarrow G(x))$, and hence $R(r)$, contradiction. So the assumption must be false, and we have proven $\forall G \left(r \neq (\hat{y})G(y)\right)$. By exploiting other Russellian antinomies, one can indeed provide any finite number of terms all of which fail to denote value-ranges; see [19].[12]

We shall come back to these matters in section 2.4. For now we note that, as Heck [16] has shown, Robinson's arithmetic **Q** can be interpreted in **H**, and so the theory is essentially undecidable. But it is unclear, though unlikely, whether much more arithmetic can be developed within the predicative fragment. We now turn briefly to an extension of Heck's predicative fragment whose consistency was doubted in [16], viz., the Δ_1^1-CA fragment.

[12] If one identifies the value-ranges with the logical objects, as I suggest in [19], these results may be reformulated as saying that **H** proves the existence of any finite number of non-logical objects. This seems correct as far as **H** goes; however, Burgess [7] points out that, if one moves on to the ramified predicative fragment, objects that are not value-ranges of any first-level concept may well be value-ranges of concepts of higher level. Similar observations pertain to the theories discussed below.)

2.3. The Δ^1_1-CA Fragment

Inspection of the proof of the Russell antinomy shows that the instance of comprehension used has complexity Σ^1_1, i.e., the comprehension formula has the form $\exists F\phi(F)$, where ϕ is predicative. Alternatively, one could use a comprehension formula of complexity Π^1_1 (that is, of the form $\forall F\phi(F)$ with predicative ϕ), viz., the variant $\forall F(x = (\hat{z})F(z) \to \neg F(x))$ of the Russell concept, to generate the contradiction. Hence both Σ^1_1- and Π^1_1-comprehension are inconsistent with schema V (in fact, either one of these schemas implies the other). This led Heck [16] to ask whether the schema of Δ^1_1-comprehension would be inconsistent with basic law V as well, where this schema can be described as the set of all instances of

$$\forall x(\phi(x) \leftrightarrow \psi(x)) \to \exists F\forall x(F(x) \leftrightarrow \phi(x)),$$

ϕ being a Σ^1_1-formula, ψ a Π^1_1-formula, and F not occurring free in ϕ. A partial answer was given in [19], to which we shall come back in the next subsection; however, the linguistic setting of the theory considered in that paper is rather different from that of **H**. In [10] it was finally shown that Δ^1_1-comprehension can indeed be consistently added to **H**. The idea of the proof is as follows: We start by building a model of the first-order fragment in the spirit of Parsons. This model has a recursively saturated elementary extension, on which we then perform the Heck construction to obtain a model of **H**. It can then be shown, by methods introduced in [2], that the recursive saturation of the first-order part forces this model to validate even Δ^1_1-comprehension.

Adding Δ^1_1-comprehension to **H** obviously results in a stronger theory; it remains doubtful, however, whether much more arithmetic can be done within this extension than within the original theory. Clearly, the observations concerning the existence of objects that are not value-ranges continue to hold for this extension of **H**.

2.4. The Theory T_Δ

To motivate the discussion in this section, let us return to the curious feature of **H** (and any consistent extension thereof) noted in subsection 2.2 above, viz., the provable existence of objects other than value-ranges. As pointed out, **H** proves $\forall F\left(r \neq (\hat{z})F(z)\right)$ for a term r that purports to denote a value range. It cannot denote the value range of any *concept*, however, because the formula from which it is derived is properly impredicative, so that no second-order entity corresponds to it. Only a VR term derived from a predicative formula will actually denote the value-range of some second-order entity. So we have many mock VR terms around—VR terms constructed out of impredicative formulae that do not define an entity in the second-order domain. As Heck [16] points out, such impredicative VR terms are mock also with regard to the membership relation: If we let $x \in y$ stand for $\exists F(y = (\hat{z})F(z) \wedge F(x))$, we will be able to prove, within **H**, that $x \in (\hat{z})\phi(z)$ always implies $\phi(x)$; however, only for predicative ϕ does **H** prove the converse.[13]

[13] On the alternative definition of membership $x \in y :\equiv \forall F(y = (\hat{z})F(z) \to F(x))$, the converse will always be provable, but the original direction only for predicative ϕ.

As shown in [19], the provable existence of non-value-ranges does not, as one might expect, hinge on the existence of mock VR terms. The theory T_Δ proved consistent there is based on a linguistic setting rather different from those of the other fragments considered so far. This is because the value range operator is construed as a third-order function symbol, which accordingly can only be attached to second-order variables, yielding terms of the form $\hat{F}$, but not to arbitrary formulae. *Grundgesetz* V is formulated as one single second-order axiom[14]:

$$\forall F \forall G\big(\hat{F} = \hat{G} \leftrightarrow \forall z(F(z) \leftrightarrow G(z))\big).$$

In addition, T_Δ has all instances of the Δ^1_1 comprehension schema as axioms. While we may of course introduce VR terms $(\hat{z})\phi(z)$ into this language by way of definitional extension, this will be possible only for formulae ϕ that are provably Δ^1_1, and so all VR terms introducible in T_Δ do actually denote value-ranges of second-order entities. Nevertheless, we can still prove that some things are not value-ranges; indeed, the non-value-ranges do not even form a concept. This can be seen as follows:

Argue in T_Δ and suppose $\exists H \forall x(H(x) \leftrightarrow \exists F(x = \hat{F}))$. Then, essentially, the Σ^1_1- and the Π^1_1-version of the Russell formula are equivalent, and hence Δ^1_1: $\exists F(x = \hat{F} \wedge \neg F(x)) \leftrightarrow \forall G(Hx \wedge (x = \hat{G} \rightarrow \neg G(x)))$. Thus, a second-order entity is defined by the Russell formula, and the antinomy follows as before. So there is no concept under which precisely the value-ranges fall. It follows that there must be objects other than value-ranges (otherwise the value range concept would simply be the universal concept, which is predicatively defined by the formula $x = x$).

3. Closing Remarks

We have surveyed a number of consistent fragments of (a second-order predicate logic reconstruction of) the theory of Frege's *Grundgesetze der Arithmetik* that arise by restricting the second-order comprehension schema. Presumably, none of these theories are of sufficient mathematical strength to provide a reconstruction of arithmetic, or even real analysis, in a Fregean spirit. But even if it turned out that a significant portion of arithmetic could be so developed, the fact remains that these theories (except the first-order fragment) prove certain claims that would seem to be unacceptable for a Fregean logicist, in particular, the existence of objects that are not value-ranges. The point is not that these theories prove the existence of infinitely many objects *simpliciter*, for, if they were to provide a foundation for arithmetic, they had better do so. But the existence of *urelemente*—in fact arbitrarily (finitely) many of them—should certainly not follow from a theory that deserves the epithet 'logical'.[15]

[14] Note that, in the context of **H** and its consistent extensions, this axiom trivially follows from schema V, but not vice versa.

[15] See, however, footnote 12 concerning ramified theories.

The fragment of section 2.3 is the strongest consistent theory obtainable from *Grundgesetze* by restricting the comprehension schema (as both Σ^1_1-CA and Π^1_1-CA lead to inconsistency), hence it seems that this strategy does not hold much promise for a Fregean foundation of mathematics. The situation may be otherwise with alternative directions of restricting or modifying Frege's system, e.g., modifications of *Grundgesetz* V (retaining full comprehension), as discussed for instance in [3].

Acknowledgement. I wish to thank an anonymous referee for a number of helpful comments.

References

[1] Anderson, David J., and Edward N. Zalta: 2004. Frege, Boolos, and Logical Objects. *Journal of Philosophical Logic* 33: 1–26.

[2] Barwise, Jon, and John Schlipf: 1975. On recursively saturated models of arithmetic. In: D. H. Saracino and V. B. Weispfenning (eds.), *Model Theory and Algebra*. Lecture Notes in Mathematics 498, Berlin: Springer-Verlag, 42–55.

[3] Boolos, George: 1986/87. Saving Frege from contradiction. *Proceedings of the Aristotelian Society* n.s. 87: 137–151.

[4] Boolos, George: 1993. Whence the contradiction? *Aristotelian Society Supplementary Volume* 67: 213–233.

[5] Boolos, George: 1997. Constructing Cantorian counterexamples. *Journal of Philosophical Logic* 26: 237–239.

[6] Burgess, John: 1998. On a consistent subsystem of Frege's *Grundgesetze*. *Notre Dame Journal of Formal Logic* 39: 274–278.

[7] Burgess, John: forthcoming. *Fixing Frege*. Princeton: Princeton University Press.

[8] Demopoulos, William (ed.): 1995. *Frege's Philosophy of Mathematics*. Cambridge, MA: Harvard University Press.

[9] Dummett, Michael: 1991. *Frege: Philosophy of Mathematics*. Cambridge, MA: Harvard University Press.

[10] Ferreira, Fernando, and Kai F. Wehmeier: 2002. On the consistency of the Δ^1_1-CA fragment of Frege's *Grundgesetze*. *Journal of Philosophical Logic* 31: 301–311.

[11] Fine, Kit: 2002. *The Limits of Abstraction*. Oxford: Oxford University Press.

[12] Frege, Gottlob: 1893. *Grundgesetze der Arithmetik, I. Band*. Jena: Hermann Pohle.

[13] Frege, Gottlob: 1903. *Grundgesetze der Arithmetik, II. Band*. Jena: Hermann Pohle.

[14] Frege, Gottlob: 1976. *Wissenschaftlicher Briefwechsel*. Edited by G. Gabriel *et al*., Hamburg: Felix Meiner.

[15] Goldfarb, Warren: 2001. First-order Frege theory is undecidable. *Journal of Philosophical Logic* 30: 613–616.

[16] Heck, Richard: 1996. The consistency of predicative fragments of Frege's *Grundgesetze der Arithmetik*. *History and Philosophy of Logic* 17: 209–220.

[17] Parsons, Terence: 1987. On the consistency of the first-order portion of Frege's logical system. *Notre Dame Journal of Formal Logic* 28: 161–168.

[18] Schroeder-Heister, Peter: 1987. A model-theoretic reconstruction of Frege's permutation argument, *Notre Dame Journal of Formal Logic* 28: 69–79.

[19] Wehmeier, Kai F.: 1999. Consistent fragments of *Grundgesetze* and the existence of non-logical objects. *Synthese* 121: 309–328.

Logic and Philosophy of Science
UC Irvine
Irvine, CA 92697-5100
USA

E-mail: wehmeier@uci.edu

On a Russellian Paradox about Propositions and Truth

Andrea Cantini

Abstract. We deal with a paradox involving the relations between propositions and sets (Appendix B of *Principles of Mathematics*), and the problem of its formalization. We first propose two (mutually incompatible) abstract theories of propositions and truth. The systems are predicatively inspired and are shown consistent by constructing suitable inductive models.

We then consider a reconstruction of a theory of truth in the context of (a consistent fragment of) Quine's set theory NF. The theory is motivated by an alternative route to the solution of the Russellian difficulty and yields an impredicative semantical system, where there exists a high degree of self-reference and yet paradoxes are blocked by restrictions to the diagonalization mechanism.

1. A Paradox of Russell Concerning the Type of Propositions

The doctrine of types is put forward tentatively in the second appendix of the *Principles* (§500). It is assumed that to each propositional function ϕ a range of significance is associated, i.e., a class of objects to which the given ϕ applies in order to produce a proposition; moreover, precisely the ranges of significance form types. However there are objects that are not ranges of significance; these are just the individuals and they form the lowest type. The next type consists of classes or ranges of individuals; then one has classes of classes of objects of the lowest type, and so on.

Once the hierarchy is accepted, new difficulties arise; in particular, if one accepts that *propositions form a type* (as they are the only objects of which it can be meaningfully asserted that they are true or false). This is a crucial point and leads Russell to a contradiction, which explicitly involves semantical notions.

First of all, since it is possible to form types of propositions, there are more types of propositions than propositions, by Cantor's argument. But then there is an argument which, if sound, apparently refutes Cantor's theorem: we can inject types of propositions into propositions by appealing to the notion of logical product.[1]

[1] A first discussion of the problem already appears in §349 of [20], where Russell deals with cases in which the conclusion of Cantor's theorem is plainly false. In this context, he mentions the crucial paradoxical embedding of classes into propositions, without solution: "I reluctantly leave the problem to the ingenuity of the reader", p. 368.

Indeed, let m be a type of propositions; to m we can associate a proposition Πm which expresses that "every proposition of m is true" (to be regarded as a possibly infinitary conjunction or logical product). Now, if m and n are (extensionally) different types, the propositions Πm and Πn must be regarded as distinct for Russell, i.e., the map $m \mapsto \Pi m$ is injective. Of course, if one were to adopt the extensional point of view, and hence equivalent propositions should be identified, no contradiction could be derived. But for Russell nobody will identify two propositions if they are simply logically equivalent; the proper equality on propositions must be much more fine-grained than logical equivalence. For instance, the proposition "every proposition which is either an m or asserts that every element of m is true, is true" is not identical with the proposition "every element of m is true" and yet the two are certainly logically equivalent.

Of course, the conflict can easily be rephrased into the form of an explicit paradox: if we accept $\{p \mid \exists m(\Pi m = p \wedge p \notin m)\} = R$ as a well-defined type,[2] we have, by injectivity of Π,

$$\Pi R \in R \Leftrightarrow \Pi R \notin R,$$

whence a contradiction.

So, if we stick to injectivity of Π, we have to change some basic tenet, e.g., to reject the assumption that propositions form one type, and hence that they ought to have various types, while logical products ought to have propositions of the same type as factors.

This will be eventually the base of the 1908 solution, but here Russell refutes the suggestion as harsh and artificial; as the reader can verify from the text,[3] he still believes that the set of all propositions is a counterexample to Cantor's theorem.

The *Principles of Mathematics*, as its coeval Fregean second volume of the *Grundgesetze*, conclude with an unsolved antinomy and Russell declares that "what the complete solution of the difficulty may be", he has not succeeded in discovering; "but as it affects the very foundations of reasoning", he earnestly commends "the study of it to the attention of all students of logic" ([20]: 528).

1.1. A Formal Outlook

In the literature there have been attempts to connect Russell's contradiction on propositions to modal paradoxes, see [17]; quite recently, Cocchiarella [5] shows how to resolve the contradiction in intensional logics that are equiconsistent with NFU,

[2] Of course, in the definition of R the quantifier ranges over types of propositions.

[3] See [20]: 527, footnote: "It might be doubted whether the relation of propositions to their logical products is one-one or many one. For example, does the logical product of p and q and r differ from that of pq and r? A reference to the definition of the logical product (p. 21) will set this doubt at rest; for the two logical products in question, though equivalent, are by non means identical. Consequently there is a one-one relation of all ranges of propositions to some propositions, which is directly contradictory to Cantor's theorem."

Quine's set theory with atoms. There is also a book of P. Grim [12], which is entirely devoted to related issues .

In the following we first show that Russell's argument can naturally be formalized and resolved within the framework of a theory PT of operations, propositions and truth, which is closely related to Aczel's (classical) Frege structures [1]. We then consider a variant of PT, proving that the very notion of propositional function defines a propositional function (this is refuted in PT).

PT will comprise the axioms of combinatory logic with extensionality [2] and the abstract axioms for truth and propositions (T, P respectively). We assume that the language contains individual constants $\dot{\rightarrow}$, $\dot{\wedge}$, $\dot{\neg}$, $\dot{\forall}$, representing the logical operations $\rightarrow$, $\wedge$, $\neg$, $\forall$, and individual constants $\doteq$, $\dot{T}$, $\dot{P}$, representing the ground predicate symbols $=$, T and P. We implicitly assume suitable independence axioms among the dotted symbols (e.g., $\dot{\neg} \neq \dot{\vee}$, $\dot{\neg}x \neq \dot{\vee}x$, etc.), granting provability of a formal analogue of the unique readability property.

It is then straightforward to define an operation $A \mapsto [A]$, which assigns to each formula of the language a term [A] with the same free variables of A, which designates the "propositional object" associated with A.[4] As usual, since we can define lambda abstraction, we can identify class abstraction $\{x \mid A\}$ with $\lambda x.[A]$.

We also define:

$$\begin{aligned}
PF(f) &\Leftrightarrow (\forall x)(P(fx));\\
\Pi f &:= \dot{\forall}(\lambda x(\dot{\rightarrow}(fx)x));\\
a \equiv_e b &\Leftrightarrow (Ta \leftrightarrow Tb);\\
a =_e b &\Leftrightarrow \forall x(T(ax) \leftrightarrow T(bx));\\
f \subseteq P &\Leftrightarrow \forall x(T(fx) \rightarrow P(x));\\
\dot{\vee}ab &:= \dot{\neg}(\dot{\wedge}(\dot{\neg}a)(\dot{\neg}b));\\
\dot{\exists} f &:= \dot{\neg}(\dot{\forall}(\lambda x.\dot{\neg}(fx))).
\end{aligned}$$

As to the terminology, if $PF(f)$ is assumed, we say that f is a *propositional function*; sometimes we use $x \in f$ instead of $T(fx)$.

The point we wish to raise is that at the very beginning of his foundational work, Russell hits upon arguments which naturally require a framework where semantical notions as well as the logical notion of set (as extension of propositional function) live on the same par.

Definition 1. We list the basic principles for propositions and truth; we essentially extend the principles implicit in the definition of *Frege structure à* la Aczel [1], with a few extra axioms.

P1 $P([x = y]) \wedge (T([x = y]) \leftrightarrow x = y)$;

P2 $T(a) \rightarrow P(a)$;

[4] This is an idea of D. Scott, see [3].

P3 $P(a) \to T([P(a)])$;

P4 $P([P(a)]) \to P(a)$;

P5 $P([T(a)]) \leftrightarrow P(a)$;

P6 $T([T(a)]) \leftrightarrow T(a)$;

P7 $P(a) \to (\neg T(a) \to T(\dot{\neg} a))$;

P8 $T(\dot{\neg} a) \to \neg T(a)$;

P9 $P(\dot{\neg} a) \leftrightarrow P(a)$;

P10 $P(a) \wedge (T(a) \to P(b)) \to P(\dot{\to} ab)$;

P11 $P(\dot{\to} ab) \to (T(a) \to P(b))$;

P12 $P(\dot{\to} ab) \to (T(a) \to T(b) \to T(\dot{\to} ab))$;

P13 $T(\dot{\to} ab) \to (T(a) \to T(b))$;

P14 $P(a) \wedge P(b) \leftrightarrow P(\dot{\wedge} ab)$;

P15 $T(\dot{\wedge} ab) \leftrightarrow T(a) \wedge T(b)$;

P16 $\forall x P(fx) \leftrightarrow P(\dot{\forall} f)$;

P17 $T(\dot{\forall} f) \leftrightarrow \forall x(T(fx))$.

Remark 1. (i) The axioms above imply a *strict interpretation* of (classically defined) disjunction and existential quantifier. By contrast with Aczel's original framework, it is assumed that it makes sense to use the predicates 'to be a proposition' and 'to be true' as logical constructors on the same par as the standard logical operators.

(ii) The question "to which kinds of object can truth be rightly attributed" is a debatable issue among philosophers. For instance, in the Tarskian approach or in Kripke's paper [16] truth is attributed to (objects representing) *sentences*, i.e., to elements of an inductively defined syntactical category. In contrast, we emphasize that here propositions form an abstract collection of objects, which are only required to meet certain broad closure conditions; being members of a combinatory structure, they can be freely combined for obtaining self-referential side effects for free. In general no induction on propositions is assumed (even if this might be true in some model, cf. §2). Also, the system is non-committal about the (delicate) question of defining a proper intensional equality for propositions. Here propositions inherit a neutral equality relation, determined by the applicative behaviour in the ground structure.

Lemma 1 (PT). (i) *If $PF(f)$ and $f \subseteq P$, then Πf is a proposition such that*

$$T(\Pi f) \leftrightarrow (\forall x(T(fx) \rightarrow T(x))).$$

Moreover:

$$\begin{aligned}
T(a) &\leftrightarrow T(\dot{\neg}(\dot{\neg}a));\\
T([\neg T(a)]) &\leftrightarrow T(\dot{\neg}a);\\
T([P(a)]) &\leftrightarrow P(a);\\
&\quad\ \exists x \neg T([\neg P(x)]);\\
P(\dot{\vee}ab) &\leftrightarrow P(a) \wedge P(b);\\
P(a) \wedge P(b) \rightarrow (T(\dot{\vee}ab) &\leftrightarrow T(a) \vee T(b));\\
P(\dot{\exists}f) &\leftrightarrow \forall x P(fx);\\
\forall x P(fx) \rightarrow (T(\dot{\exists}f) &\leftrightarrow \exists x T(fx)).
\end{aligned}$$

(ii) *We also have:*

$$\begin{aligned}
\Pi f = \Pi g &\rightarrow f = g\\
PF(a) \wedge PF(b) \wedge a =_e b &\rightarrow (\Pi a \equiv_e \Pi b)
\end{aligned}$$

(iii) Π *is not extensionally injective, i.e., there exist propositional functions a, b such that*

$$\Pi a \equiv_e \Pi b \wedge \neg(a =_e b).$$

Proof. As to (i), we only check the first claim (using axioms for $\dot{\rightarrow}$, $\dot{\forall}$)

$$\begin{aligned}
PF(f) \wedge f \subseteq P &\Rightarrow P(fx) \wedge (T(fx) \rightarrow P(x))\\
&\Rightarrow P(\dot{\rightarrow}(fx)x)\\
&\Rightarrow PF(\lambda x.\dot{\rightarrow}(fx)x)\\
&\Rightarrow P(\Pi f)\\
&\Rightarrow T(\Pi f) \leftrightarrow \forall x(T(fx) \rightarrow T(x))
\end{aligned}$$

(ii): apply injectivity of $\dot{\forall}$, $\dot{\rightarrow}$ and extensionality for operations.

(iii): choose $a = \{[K = K], [S = S]\}$[5] and $b = \{[K = K]\}$; then $\Pi a \equiv_e \Pi b$, but a and b are extensionally distinct propositional functions). □

Clearly, we can derive the Tarskian T-schema:

Proposition 1. *If A is an arbitrary formula, then* PT *proves:*

$$P([A]) \rightarrow (T([A]) \leftrightarrow A)$$

[5] Of course $\{a, b\}$ stands for $\{x | x = a \vee x = b\}$.

Proposition 2 (Russell's Appendix B, [20]). *The term*

$$\{x \mid \exists m(PF(m) \wedge m \subseteq P \wedge x \notin m \wedge x = \Pi m)\}$$

does not define a propositional function, provably in PT.

Proof. Let

$$R := \{x \mid \exists m(PF(m) \wedge m \subseteq P \wedge x \notin m \wedge x = \Pi m)\}.$$

Assume by contradiction that $PF(R)$. Then, by applying the closure conditions of P, T and lemma 1:

$$(\forall x \in R)(P(x)).$$

Hence:

$$P(\Pi R).$$

Now we have, for some m:

$$\Pi R \in R \Rightarrow \Pi R = \Pi m \wedge PF(m) \wedge m \subseteq P \wedge \Pi R \notin m.$$

By the previous lemma (Π is 1-1), $R = m$ and hence $\Pi R \notin R$. But $\Pi R \notin R$ implies $\Pi R \in R$, since R is a propositional function of propositions. □

Lemma 2 (PT). *If P is a propositional function, then PF itself is a propositional function.*

Proof. If P is a propositional function, we have:

$$\begin{aligned}
&\Rightarrow \forall x.P([P(x)]) \\
&\Rightarrow \forall x.P([P(fx)]) \\
&\Rightarrow \forall f P([\forall x.P(fx)]) \\
&\Rightarrow \forall f P([PF(f)])
\end{aligned}$$

Theorem 1 (PT). *PF and P are not propositional functions.*

Proof. By the previous lemma, it is enough to show that PF is not a propositional function.

Assume PF is a propositional function. The axioms on the relation between P and $\forall$ imply:

$$\forall x \forall u.P([P(xu)]).$$

Hence using the axiom $P([P(xu)]) \rightarrow P(xu)$, we conclude:

$$\forall x.PF(x),$$

against the previous proposition. □

Alternative argument: If $\lambda x.PF(x)$ is a propositional function, we can show with the implication axioms that

$$W := \{x \mid PF(x) \wedge (PF(x) \to \dot{\neg}(xx))\}$$

is a propositional function. Then the standard Russell paradox arises.

The conclusion is that under relatively mild hypotheses (a notion of truth which obeys to classical laws, endowed with an abstract notion of proposition) the paradox disappears. Clearly, no propositional function reasonably defining the power class of the collection of propositions can exist in the above framework; on the same par, the collection of propositions cannot give rise to a well-defined propositional function. The solution is compatible with Russell's no-class theory: it could be assumed that the universe of classes exactly includes those collections which are represented by terms of the form $\{x \mid A(x)\}$, where $A(x)$ defines a propositional function, cf. the model construction of §2.

1.2. An Alternative Theory of Truth and Propositions

Definition 2. We describe a variant AT of the system PT, which is so devised that the assumption "$PF(x)$ is a proposition" is consistent:

A1 $P([x = y]) \wedge (T([x = y]) \leftrightarrow x = y)$;

A2 $T(a) \to P(a)$;

A3 $\forall x T([\neg P(x)])$;

A4 $P([T(a)]) \leftrightarrow P(a)$;

A5 $T([T(a)]) \leftrightarrow T(a)$;

A6 $P(a) \to (\neg T(a) \to T(\dot{\neg}a))$;

A7 $T(\dot{\neg}a) \to \neg T(a)$;

A8 $P(\dot{\neg}a) \leftrightarrow P(a)$;

A9 $P(a) \wedge P(b) \leftrightarrow P(\dot{\wedge}ab)$;

A10 $T(\dot{\wedge}ab) \leftrightarrow T(a) \wedge T(b)$;

A11 $\forall x P(fx) \leftrightarrow P(\dot{\forall}f)$;

A12 $T(\dot{\forall}f) \leftrightarrow \forall x(T(fx))$.

The typical principle of the system is A3, according to which no claim about $P(a)$ can be internally true. A3 implies with A2, A8 the following principle (Λ):

$$\forall x.P([P(x)]).$$

It follows from (Λ) that AT is inconsistent with the axiom:

$$P([P(x)]) \to P(x).$$

Indeed, let L be the Liar object $L = \dot{\neg}(L)$. Since $P([P(L)])$ holds, we have $P(L)$ and we can conclude with the negation axioms that $T(\dot{\neg}L) \leftrightarrow \neg T(L)$, whence a contradiction.

Moreover, if we apply (Λ) and A11, we obtain:

Proposition 3 (AT). *$\lambda x.PF(x)$ is a propositional function.*

Observe also that A3 implies that the PT-axiom $P(a) \to T([P(a)])$ is inconsistent with AT (choose $a := [x = y]$).

On the other hand, similarly to PT, we can prove in AT:

$$T([\neg T(a)]) \leftrightarrow T(\dot{\neg}a).$$

In order to appreciate the difference in strength between AT and PT, it may be of interest to compare the closure properties of propositional functions in the two systems. First of all, both systems are closed under *elementary comprehension*. Indeed, if $\vec{a}$ is a finite list $a_0, \ldots, a_n$ of variables distinct from x, we say that $A(x, \vec{a})$, where $FV(A) \subseteq \vec{a}, x$,[6] is *elementary in* $\vec{a}$, if A is inductively generated from atomic formulas of the form $t = s$, $t \in a_i$ (with $0 \leq i \leq n$ and no variable of $\vec{a}$ free in t, s) by means of $\neg$, $\wedge$ and quantification on variables not occurring in the list $\vec{a}$.

Let PT$\cap$AT be the theory consisting of those axioms which are common to PT and AT.

Proposition 4 (PT$\cap$AT). *If $A(x, \vec{a})$ is elementary in $\vec{a}$ and each a_i of the list $\vec{a}$ is a propositional function, then $\{x \mid A(x, \vec{a})\}$ is a propositional function such that*

$$\forall u(u \in \{x \mid A(x, \vec{a})\} \leftrightarrow A(u, \vec{a})).$$

However, while propositional functions are closed under disjoint union (or join) in PT, the same cannot be proved in AT.

Let us justify this claim. We introduce a kind of *sequential conjunction* $\odot$, due to Aczel [1]:

$$a \odot b := a \mathbin{\dot{\wedge}} (a \to b).$$

Clearly PT proves:

$$P(a) \wedge (T(a) \to P(b)) \to P(a \odot b) \wedge T(a \odot b) \leftrightarrow T(a) \wedge T(b).^{7}$$

[6] $FV(E)$ is the set of free variables occurring in the expression E.

Since we have combinatory logic as underlying theory, we can assume to have an ordered pairing operation $x, y \mapsto (x, y)$ with projections $u \mapsto u_0, u \mapsto u_1$. Therefore it makes sense to define

$$\Sigma(a, f) := \{u \mid (u = (u_0, u_1) \wedge u_0 \in a \odot u_1 \in f(u)_0)\}.$$

Proposition 5. (i) PT *proves that, if a is a propositional function and fx is a propositional function whenever $x \in a$, then $\Sigma(a, f)$ is a propositional function satisfying* (J):

$$u \in \Sigma(a, f) \leftrightarrow u = (u_0, u_1) \wedge u_0 \in a \wedge u_1 \in f(u)_0.$$

(ii) AT *proves the weak power class axiom: for every propositional function a, there exists a propositional function $Pow^-(a)$ such that*

$$\forall u(u \in Pow^-(a) \rightarrow PF(u) \wedge u \subseteq a);$$
$$\forall u(PF(u) \wedge u \subseteq a \rightarrow \exists b(PF(b) \wedge b \in Pow^-(a) \wedge u =_e b)).$$

As to the proof, (i) is an immediate consequence of the definition of $\odot$, while (ii) essentially depends on the fact that $P(x)$ is a proposition for every x, once we set

$$Pow^-(a) = \{u \mid PF(u) \wedge \exists b(PF(b) \wedge u = b \cap a)\}.$$

Corollary 1. AT+(J) *is inconsistent.*

This follows by adapting Russell's paradox along the lines of Feferman, see [7].

1.3. PT and AT Are Non-Cantorian

The claim means that in neither truth theory there is a good analogue of the power set in terms of propositional functions. The reason is given by the results below, which are provable in the common subtheory of PT and AT (so they do not involve properties of strong implication nor the fact that $\lambda x.PF(x)$ is a propositional function).

Definition 3.

(i) (A formula of our language) $\varphi(x)$ is called *extensional* iff $\forall f \forall g(PF(f) \wedge PF(g) \wedge f =_e g \wedge \varphi(f) \rightarrow \varphi(g))$;

(ii) a formula $\varphi(x)$ such that $\forall x(\varphi(x) \rightarrow PF(x))$ is *non-trivial*, provided there are propositional functions x, y such that $\varphi(x), \neg\varphi(y)$.

[7] We assume that $\leftrightarrow$ binds more strongly than $\wedge$.

Lemma 3 ("Inseparability", in PT∩AT). *Let φ_1, φ_2 be extensional formulas and assume that there exist propositional functions x_1, x_2 such that $\varphi_1(x_1)$, $\varphi_2(x_2)$. Then φ_1 and φ_2 are PF-inseparable, i.e., for no propositional function x_3 we can have:*

$$\forall u(PF(u) \wedge \varphi_2(u) \to u \notin x_3) \wedge \forall u(PF(u) \wedge \varphi_1(u) \to u \in x_3)$$

Proof. Assume that x_3 is a propositional function such that

$$\forall u(PF(u) \wedge \varphi_2(u) \to u \notin x_3).$$

It is enough to produce a propositional function $g := g(x_1, x_2, x_3)$ such that

$$g \notin x_3 \wedge \varphi_1(g).$$

By the fixed point lemma, choose an element g such that $g = Gg$, where

$$Gh = \{u \mid (u \in x_1 \wedge h \notin x_3) \vee (u \in x_2 \wedge h \in x_3)\}.$$

Then, using the common axioms on $\wedge$, $\neg$ and the assumption that x_3, x_2, x_1 are propositional functions, we have that g is a propositional function.

If $g \in x_3$, then $g =_e x_2$. Since $\varphi_2(x_2)$, also $\varphi_2(g)$; thus $g \notin x_3$.

Hence $g \notin x_3$, which yields $g =_e x_1$. But $\varphi_1(x_1)$; so $\varphi_1(g)$ by extensionality. □

Theorem 2 (PT∩AT). *No propositional function f can be both extensional and non-trivial.*

So, for instance, there cannot exist a propositional function playing the role of the power set of $\{\emptyset\}$, i.e., whose "range of significance" is exactly the collection of all propositional functions $\subseteq \{\emptyset\}$: for this would be non-trivial and extensional.

The results above are extensions to the context of Frege structures of results holding for theories of explicit mathematics, see [4].

2. About Models

We outline two inductive model constructions for PT and AT, respectively.

2.1. PT-Models

The basic idea is to consider any given combinatory algebra as ground universe and to produce, by generalized inductive definition, the collections of propositions and truths. This cannot be simply rephrased as a standard monotone inductive definition, because the clause for introducing a proposition of implicative form makes use (negatively) of the collection of truths. Nevertheless, we can adapt a trick of Aczel [1].[8]

[8] As an alternative, we can define the model by transfinite recursion over the ordinals, in much the same way as the model for Feferman's theories with the so-called join axiom.

Fix an extensional combinatory algebra $\mathcal{M}$; let $|M|$ be the universe of $\mathcal{M}$. If $X = \langle X_0, X_1\rangle$, $Y = \langle Y_0, Y_1\rangle$, and X_0, X_1, Y_0, Y_1 are subsets of M, define

$$X \leq Y \Leftrightarrow X_0 \subseteq Y_0 \wedge (\forall a \in X_0)(a \in X_1 \leftrightarrow a \in Y_1).$$

Let $\mathcal{F}$ be the family of all pairs $X = \langle X_0, X_1\rangle$ of subsets of $|M|$, satisfying the restriction $X_1 \subseteq X_0$. If $X \in \mathcal{F}$, we call X *suitable*.[9]

Lemma 4. *The structure $\langle \mathcal{F}, \leq\rangle$ is a complete partial ordering.*[10]

We then define an operator Γ on suitable subsets of $|M|$. $\Gamma(X)$ can be given by specifying two operators $\Gamma_0(X), \Gamma_1(X)$ such that $\Gamma(X) = \langle \Gamma_0(X), \Gamma_1(X)\rangle$. $\Gamma_0(X)$ is defined by the following formula $A_0(x, X)$:

$$\begin{aligned}\exists u \exists v\ [\ & (x = [u = v]) \vee (x = [Pu] \wedge u \in X_0) \vee \\ & \vee\ (x = [Tu] \wedge u \in X_0) \vee \\ & \vee\ (x = (\dot{\neg} u) \wedge u \in X_0) \vee \\ & \vee\ ((x = (u \dot{\vee} v)\ \vee\ x = (u \dot{\wedge} v)) \wedge u \in X_0 \wedge v \in X_0) \vee \\ & \vee\ (x = (u \dot{\rightarrow} v)\ \wedge\ u \in X_0\ \wedge\ (u \notin X_1 \vee v \in X_0)) \vee \\ & \vee\ ((x = \dot{\forall} u \vee x = \dot{\exists} u) \wedge \forall y (uy \in X_0))].\end{aligned}$$

$\Gamma_1(X)$ is defined by the following formula $A_1(x, X)$:

$$\begin{aligned}\exists u \exists v\ [\ & (x = [u = v] \wedge u = v) \vee \\ & \vee\ (x = [Pu] \wedge u \in X_0) \vee (x = [Tu] \wedge u \in X_1) \vee \\ & \vee\ (x = (\dot{\neg} u) \wedge u \in X_0 \wedge u \notin X_1) \vee \\ & \vee\ (x = (u \dot{\vee} v) \wedge u \in X_0 \wedge v \in X_0 \wedge (u \in X_1 \vee v \in X_1)) \vee \\ & \vee\ (x = (u \dot{\wedge} v) \wedge u \in X_0 \wedge v \in X_0 \wedge u \in X_1 \wedge v \in X_1) \vee \\ & \vee\ (x = (u \dot{\rightarrow} v)\ \wedge \\ & \quad \wedge (u \in X_0\ \wedge\ (u \notin X_1 \vee v \in X_0) \wedge (u \notin X_1 \vee v \in X_1))) \vee \\ & \vee\ (x = \dot{\forall} u \wedge \forall y (uy \in X_0) \wedge \forall y (uy \in X_1)) \vee \\ & \vee\ (x = \dot{\exists} u \wedge \forall y (uy \in X_0) \wedge \exists y (uy \in X_1))].\end{aligned}$$

The following properties are immediate:

Lemma 5. (i) $\Gamma : \mathcal{F} \to \mathcal{F}$.

(ii) *$X \leq Y$, then $\Gamma(X) \leq \Gamma(Y)$ (where $X, Y \in \mathcal{F}$). Hence there are sets $X \in \mathcal{F}$, such that $X = \Gamma(X)$.*

[9] Intuitively, X_0 is a candidate for the set of propositions, while X_1 represents the corresponding notion of truth.

[10] I.e., a partial ordering in which every $\leq$-increasing sequence of elements of $\mathcal{F}$ has a least upper bound with respect to $\leq$.

Theorem 3. *If $X \in \mathcal{F}$ and $X = \Gamma(X)$, then*

$$\langle \mathcal{M}, X \rangle \models \mathrm{PT}.$$

Hence PT *is consistent.*

The proof of the theorem is straightforward.

2.2. AT-Models

As to the system AT, we first inductively define the set of propositional objects over a given combinatory algebra. We then exploit the stages assigned to propositional objects for generating the truth set.

Formally, we fix an extensional combinatory algebra $\mathcal{M}$ with universe $|M|$. We also assume that our language includes names for objects of $|M|$ (for which we adopt the same symbols). If t is a term of the expanded language, t^M stands for the value of t in $\mathcal{M}$. We are now ready to define, by transfinite recursion on ordinals, a sequence $\{\mathcal{P}_\alpha\}$ of subsets of $|M|$:

- Initial clause:

$$\mathcal{P}_0 = \{[a = b] \mid a,\ b \in M\} \cup \{[Pa] \mid a \in M\};$$

- Limit clause: if λ is a limit ordinal,

$$\mathcal{P}_\lambda = \cup\{\mathcal{P}_\alpha \mid \alpha < \lambda\};$$

- $\neg$- and $\wedge$-rules:

$$\frac{a \in \mathcal{P}_\alpha}{(\dot{\neg} a)^M \in \mathcal{P}_{\alpha+1}} \qquad \frac{a \in \mathcal{P}_\alpha \quad b \in \mathcal{P}_\alpha}{(a \dot{\wedge} b)^M \in \mathcal{P}_{\alpha+1}};$$

- $\forall$- and T-rules:

$$\frac{\text{for all } c \in |M|,\ (fc)^M \in \mathcal{P}_\alpha}{\dot{\forall} f^M \in \mathcal{P}_{\alpha+1}} \qquad \frac{a \in \mathcal{P}_\alpha}{(\dot{T} a)^M \in \mathcal{P}_{\alpha+1}}.$$

In a similar way, we recursively produce a sequence $\{\mathcal{T}_\alpha\}$ of subsets of $|M|$, approximating the truth set :

- Initial clause:

$$\frac{\mathcal{M} \models a = b}{[a = b]^M \in \mathcal{T}_0};$$

- Limit clause: if λ is a limit ordinal,

$$\mathcal{T}_\lambda = \cup\{\mathcal{T}_\alpha \mid \alpha < \lambda\};$$

- First successor clause: if $a \in \mathcal{P}_\alpha$,

$$a \in \mathcal{T}_{\alpha+1} \Leftrightarrow a \in \mathcal{T}_\alpha;$$

- Second successor clause: assume $a \in \mathcal{P}_{\alpha+1} - \mathcal{P}_\alpha$. We distinguish several cases according to the form of a, i.e., $a = \dot{\forall} f$, $\dot{T} b$, $\dot{\neg} b$, $b \dot{\wedge} c$ (respectively).

  1. $\forall$- and T-clauses:

$$\frac{\text{for all } c \in |M|,\ (fc)^M \in \mathcal{T}_\alpha}{\dot{\forall} f^M \in \mathcal{T}_{\alpha+1}} \qquad \frac{b \in \mathcal{T}_\alpha}{(\dot{T} b)^M \in \mathcal{T}_{\alpha+1}};$$

  2. $\wedge$- and $\neg$-clauses:

$$\frac{b \in \mathcal{T}_\alpha \quad c \in \mathcal{T}_\alpha}{(b \dot{\wedge} c)^M \in \mathcal{T}_{\alpha+1}} \qquad \frac{b \notin \mathcal{T}_\alpha}{(\dot{\neg} b)^M \in \mathcal{T}_{\alpha+1}}.$$

By transfinite induction on ordinals, it is not difficult to verify:

Lemma 6. *If $a \in M$, then*

$$\begin{aligned} a \in \mathcal{T}_\alpha &\Rightarrow a \in \mathcal{P}_\alpha; \\ \alpha \leq \beta &\Rightarrow \mathcal{P}_\alpha \subseteq \mathcal{P}_\beta \wedge \mathcal{T}_\alpha \subseteq \mathcal{T}_\beta; \\ a \in \mathcal{P}_\alpha &\Rightarrow a \in \mathcal{T}_\alpha \vee (\dot{\neg} a) \in \mathcal{T}_{\alpha+1}; \\ (\dot{\neg} a) \in \mathcal{T}_\alpha &\Rightarrow a \notin \mathcal{T}_\beta \quad (\beta \textit{ arbitrary}). \end{aligned}$$

We choose $\mathcal{P} = \cup\{\mathcal{P}_\alpha \mid \alpha \text{ ordinal}\}$, $\mathcal{T} = \cup\{\mathcal{T}_\alpha \mid \alpha \text{ ordinal}\}$; of course, $\mathcal{P}$, $\mathcal{T}$ depend on the underlying combinatory algebra $\mathcal{M}$, but we leave this fact implicit.

Lemma 7. *Let $\mathcal{O}$ be the open term model of combinatory logic plus extensionality.*[11] *Then it holds over $\mathcal{O}$:*

$$\mathcal{P} = \mathcal{P}_\omega \textit{ and } \mathcal{T} = \mathcal{T}_\omega.$$

Proof. Assume that we have proved

$$\mathcal{P} = \mathcal{P}_\omega. \tag{1}$$

By lemma 6, if $a \in \mathcal{T}_\delta$, then $a \in \mathcal{P}_\delta$. By assumption, $a \in \mathcal{P}_k$, for some finite k. By the third claim of lemma 6, either $a \in \mathcal{T}_k$ or $(\dot{\neg} a) \in \mathcal{T}_{k+1}$. In the first case we

[11] For details, see [2].

are done; the second case implies $a \notin \mathcal{T}_\delta$ (again lemma 6, last claim), contradiction. Hence $\mathcal{T}_\delta \subseteq \mathcal{T}_\omega$. But the converse inclusion trivially holds and hence $\mathcal{T} = \mathcal{T}_\omega$.

It remains to check (1). It is sufficient to define a recursively enumerable derivability relation $\vdash$ over the term model, such that, for every $a \in \mathcal{O}$,

$$\vdash a \Leftrightarrow a \in \mathcal{P}.$$

But this is straightforward: the axioms of $\vdash$ will have the form $[t = s]$, $[P(t)]$, while the inference rules correspond to the positive inductive clauses generating the sequence $\{\mathcal{P}_\alpha\}$. Of course, the clause for $\forall$ can be rephrased as a finitary inference: from $\vdash ax$, infer $\vdash \dot{\forall} a$, provided x is not free in a.

It is then easy to check that the derivability relation is closed under substitution, that is, for arbitrary terms a, s:

$$\vdash a(x) \Rightarrow \vdash a[x := s].$$

This property together with the fact that $\mathcal{O}$ is the open term model readily yields the initial claim (1) (proofs are carried out by induction on the definition of $\vdash$ and by transfinite induction on ordinals). □

Theorem 4. $\langle \mathcal{M}, \mathcal{P}, \mathcal{T} \rangle \models \mathrm{AT}$.

Proof-theoretic digression. Of course, we can consider applied versions of PT and AT. Indeed, let PTN (ATN) be PT (AT) extended with a predicate N for the set of natural numbers, constants for 0, successor, predecessor and conditional on N, the induction schema for natural numbers for N. Then:

Theorem 5. (i) PTN *is proof-theoretically equivalent to ramified analysis of arbitrary level below* ε_0.

(ii) *If* N*-induction is restricted to propositional functions, the resulting system* PTN_c *is proof-theoretically equivalent to Peano arithmetic*.

The proof follows well-known paths; the lower bound can be obtained by embedding in PTN a system of the required strength, for instance Feferman's EM_0+J [7].

As to the upper bound, it is possible to provide a proof-theoretic analysis of PTN with predicative methods (partial cut elimination and asymmetric interpretation into a ramified system with levels $< \varepsilon_0$). A quick proof of the conservation result exploits recursively saturated models.

Concerning the strength of ATN, we do not have a definite result yet, but we believe that the following is true:

Conjecture 6. (i) ATN has the same proof-theoretic strength as ACA, the system of second order arithmetic based on arithmetical comprehension.

(ii) ATN with number theoretic induction restricted to propositional functions is proof-theoretically equivalent to Peano arithmetic.

As to possible routes for proving (i)–(ii), one ought to consider lemma 7 and the methods of Glass [10].

3. Stratified Truth?

We now explore an alternative route, which takes into account the possibility of dealing with the paradox in a fully impredicative, extensional framework, Quine's set theory NF. In the new model, the set of all propositions and the set of all truths do exist, and, to a certain extent, the notion of truth has rather strong closure properties.

We first describe the formal details. $\mathcal{L}_s$ is the elementary set theoretic language, which comprises the binary predicate symbol $\in$. $\mathcal{L}_s$-terms are simply individual variables ($x, y, z, \ldots$) and prime formulas (atoms) have the form $t \in s$ (t, s terms). $\mathcal{L}_s$-formulas are inductively generated from prime formulas by means of sentential connectives and quantifiers. The elementary set theoretic language $\mathcal{L}_s^+$ is obtained by adding to $\mathcal{L}_s$ the abstraction operator $\{- \mid -\}$; $\mathcal{L}_s^+$-terms and formulas are then simultaneously generated. The clause for introducing class terms has the form: if φ is a formula, then $\{x \mid \varphi\}$ is a term where $FV(\{x \mid \varphi\}) = FV(\varphi) - \{x\}$.

As usual, two terms (formulas) are called α-congruent if they only differ by renaming bound variables; we identify α-congruent terms (formulas).

3.1. Stratified Comprehension

As usual for Quine's systems, we need the technical device of *stratification*; we also define a restricted notion thereof, which is motivated by the consideration of "loosely predicative" class existence axioms.

(i) φ is *stratified* iff it is possible to assign a natural number (type in short) to each term occurrence[12] of φ in such a way that

1. if $t \in s$ is a subformula of φ, the type of s is one greater than the type of t;
2. all free occurrences of the same variable in any subformula of φ have the same type;
3. if x is free in ψ and $\forall x \psi$ is a subformula of φ, then the 'x' in $\forall x$ and the free occurrences of x in ψ receive the same type;
4. if $t = \{x \mid \beta\}$ occurs in φ, x is free in β, then t is assigned a type one greater than the type assigned to x, and all the free occurrences of x in β receive the same type.

(ii) $\{x \mid \varphi\}$ is stratified if φ is stratified;

[12] Individual constants included; these can be given any type compatible with the clauses below.

(iii) a stratified term $\{x \mid \varphi(x, \vec{y})\}$ is *loosely predicative* iff for some type $i \in \omega$, $\{x \mid \varphi(x, \vec{y})\}$ has type $i+1$, no (free or bound) variable of $\varphi(x, \vec{y})$ is assigned type greater than $i+1$; a stratified term $\{x \mid \varphi(x, \vec{y})\}$ is *predicative* iff $\{x \mid \varphi(x, \vec{y})\}$ is loosely predicative and in addition no quantified variable of $\varphi(x, \vec{y})$ is assigned the same type as $\{x \mid \varphi(x, \vec{y})\}$ itself.

(iv) φ is $n+1$-*stratified* iff φ is stratified by means of $0, \ldots, n$.

For instance, $\bigcup a = \{x \mid (\exists y \in a)(x \in y)\}$ is not loosely predicative, since it requires type 2, but $\bigcup a$ itself has type 1; $a \cap b = \{x \mid x \in a \wedge x \in b\}$ is predicative.

Definition 4. The system NF comprises:

1. predicate logic for the extended language;[13]

2. class extensionality: $\forall x \forall y(x =_e y \to x = y)$, where

$$t = s :\Leftrightarrow \forall z(t \in z \to s \in z);$$
$$t =_e s :\Leftrightarrow \forall x(x \in t \leftrightarrow x \in s);$$

3. stratified explicit comprehension SCA: if φ is stratified, then

$$\forall u(u \in \{x \mid \varphi(x, \vec{y})\} \leftrightarrow \varphi(u, \vec{y})).$$

Other systems.

(i) NFP (NFI) is the subsystem of NF, where SCA is restricted to (loosely) predicative abstracts.

(ii) NF_k (NFI_k, NFP_k) is the subsystem of NF (NFI, NFP_k), where (at most) k types are allowed for stratification.

Remark 2. If Union is the axiom "$\bigcup a$ exists, for all a", then NFP+Union is equivalent to the full NF; cf. [6].

[13] To be more accurate, if the abstraction operator is assumed as primitive, it is convenient to include in the extended logic the schema:

$$\forall u(\varphi(u) \leftrightarrow \psi(u)) \to \{x \mid \varphi(x)\} = \{x \mid \psi(x)\}.$$

This would ensure that $\{x \mid \neg x = x\} = \{x \mid x \in x \wedge \neg x \in x\}$. An alternative route would be to extend the logic with a description operator, say in the style of [15], ch. VII. If this choice is adopted, the previous schema becomes provable. Be this as it may, the resulting theories are conservative over NF as formalized in the pure set theoretic language $\mathcal{L}_s$, and we won't bother the reader with further details.

3.2. Consistent NF-Subtheories

By a theorem of Crabbè [6], NFI is provably consistent in third order arithmetic. The details of the (different) consistency proofs for NFI can be found in [6] and [14]. The main idea of [6]: 133–135 is to exploit the reducibility of NFI to its fragment NFI_4; then NFI_4 is interpreted into a corresponding type theory up to level 4, TI_4, plus Amb (= the so-called schema of typical ambiguity, see [22]).[14] These steps are finitary and adapt well-known theorems of Grishin and Specker. TI_4+Amb is shown consistent by means of the Hauptsatz for second order logic (which is equivalent in primitive recursive arithmetic to the 1-consistency of full second order arithmetic, e.g., see [11]: 280). Hence:

Theorem 7. NFI *is consistent* (*in primitive recursive arithmetic plus the* 1*-consistency of second order arithmetic*).

In order to carry out a Kripke-like construction in the NF-systems and to represent the syntax, we shall essentially exploit Quine's homogeneous pairing operation, which *does require extensionality* and the existence of a copy of the natural numbers. But it is not difficult to check that Quine's pairing is indeed well-defined already in NFI.

This requires two steps. First of all, the collection of Fregean natural numbers is a set in NFI. Define:

$$\begin{aligned} \emptyset &= \{x \mid x \neq x\};\\ V &= \{x \mid x = x\};\\ 0 &= \{\emptyset\};\\ a+1 &= \{x \cup \{y\} \mid x \in a \wedge y \notin x\};\\ Cl_N(y) &\Leftrightarrow 0 \in y \wedge \forall x(x \in y \rightarrow (x+1) \in y);\\ \mathcal{N} &= \{x \mid \forall y(Cl_N(y) \rightarrow x \in y)\}. \end{aligned}$$

Then NFI grants the existence of $\mathcal{N}$; in fact, by inspection, all the above sets above are loosely predicative. Furthermore, we have, provably in NFI:

Lemma 8 (NFI).

$$Cl_N(\{x \mid \varphi(x)\}) \rightarrow \mathcal{N} \subseteq \{x \mid \varphi(x)\}; \tag{2}$$

$$(\forall x)(x \in \mathcal{N} \leftrightarrow x = 0 \vee (\exists y \in \mathcal{N})(x = y+1)); \tag{3}$$

$$\emptyset \notin \mathcal{N} \wedge (\forall x \in \mathcal{N})(V \notin x); \tag{4}$$

$$(\forall x \in \mathcal{N})(x+1 \neq 0); \tag{5}$$

$$(\forall x \in \mathcal{N})(\forall y \in \mathcal{N})(x+1 = y+1 \rightarrow x = y). \tag{6}$$

(In (2) $\{x \mid \varphi(x)\}$ must be loosely predicative.)

[14] In TI_4, $\{x^i \mid \varphi\}$ exists, provided $i = 0, 1, 2$ and φ contains free or bound variables of type $i+1$ at most.

Clearly $\mathcal{N}$ is infinite by (4) above.[15] As to the proof, (4) holds in NFI+Union, as NFI+Union $\equiv$ NF, and NF proves (4) according to a famous result of Specker [21]. On the other hand, NFI + $\neg$Union implies (4), by [6]. The claims (3), (2) with the Peano axioms are provable in NFI ((6) requires the second part of (4)).

Definition 5 (Homogeneous pairing; [19]).

$$\begin{aligned}
\phi(a) &= \{y \mid y \in a \wedge y \notin \mathcal{N}\} \cup \{y+1 \mid y \in a \wedge y \in \mathcal{N}\};\\
\theta_1(a) &= \{\phi(x) \mid x \in a\};\\
\theta_2(a) &= \{\phi(x) \cup \{0\} \mid x \in a\};\\
(a,b) &= \theta_1(a) \cup \theta_2(b);\\
Q_1(a) &= \{z \mid \phi(z) \in a\};\\
Q_2(a) &= \{z \mid \phi(z) \cup \{0\} \in a\}.
\end{aligned}$$

The definitions above are (at most) loosely predicative and hence the universe of sets is closed under the corresponding operations, provably in NFI.

Lemma 9 (NFI).

1. $\phi(a) = \phi(b) \to a = b$;
2. $0 \notin \phi(a)$;
3. $\theta_i(a) = \theta_i(b) \to a = b$, *where* $i = 1, 2$;
4. $Q_i((x_1, x_2)) = x_i$, *where* $i = 1, 2$;
5. $(x, y) = (u, v) \to x = u \wedge y = v$;
6. *the map* $x, y \longmapsto (x, y)$ *is surjective and* $\subseteq$*-monotone in each variable.*

The proof hinges upon the properties of $\mathcal{N}$ and the successor operation [19]. In particular, we below exploit the fact that Quine's pairing operation is $\subseteq$-monotone in both arguments. This is seen by inspection: the definition of (a, b) is positive in a, b.[16]

Lemma 10 (Fixed point). *Let* $A(x, a)$ *be a formula which is positive in* a. *Assume that*

$$\Gamma_A(a) = \{x \mid A(x, a)\}$$

is loosely predicative, where x, a *are given types* i, $i+1$ *respectively. Then* NFI *proves the existence of a set* c *of type* $i+1$, *such that*:

[15] Indeed, if $\emptyset \notin \mathcal{N}$, then no natural number à la Frege-Russell reduces to the empty set, i.e., there exist arbitrary large finite sets.

[16] We recall that a formula $A(x, a)$ is positive in a if every free occurrence of a in the negation normal form of A is located in atoms of the form $t \in a$, which are prefixed by an even number of negations and where $a \notin FV(t)$.

- $\Gamma_A(c) \subseteq c$;
- $\Gamma_A(a) \subseteq a \Rightarrow c \subseteq a$.

The proof is standard: observe that the set

$$c := \{x \mid \forall d(\Gamma_A(d) \subseteq d \rightarrow x \in d)\}$$

is loosely predicative.[17]

Definition 6. NFI(pair) (NFP(pair), NF(pair)) is the theory, which extends NFI (NFP, NF respectively) with a new binary function symbol $(-, -)$ for ordered pairing and the corresponding new axiom:

$$\forall x \forall y \forall u \forall v((x, y) = (u, v) \rightarrow x = u \wedge y = v)$$

It is understood that the stratification condition is lifted to the new language by stipulating that t, s in (t, s) receive the same type.

The strategic role of homogeneous pairing is clarified by two equivalence results. Below let $\mathcal{S}_1 \equiv \mathcal{S}_2$ denote the relation that holds between two formal theories $\mathcal{S}_1, \mathcal{S}_2$ whenever they are mutually interpretable.

Proposition 6. $\mathrm{NF} \equiv \mathrm{NF}_3(\mathrm{pair})$.

The proof of this statement was independently suggested by Antonelli and Holmes. The basic observation is that $\mathrm{NFP}_3(\mathrm{pair})$ proves the existence of the set E, where $\mathcal{E} := \exists y(y = E)$ and $E = \{\{\{x\}, y\} \mid x \in y\}$. But by Grishin's [13], $\mathrm{NF} \equiv \mathrm{NF}_3 + \mathcal{E}$.

We can readily extend the proposition to the subsystems NFI, NFP.

Proposition 7. $\mathrm{NFP} \equiv \mathrm{NFP}_3(\mathrm{pair})$ and $\mathrm{NFI} \equiv \mathrm{NFI}_3(\mathrm{pair})$.

Proof. Let

$$\mathcal{I} \Leftrightarrow \exists y(y = \{u \mid \cap u \neq \emptyset\}).$$

By a theorem of Grishin and [6], $\mathrm{NFI} \equiv \mathrm{NFI}_3 + \mathcal{I}$, (respectively $\mathrm{NFP} \equiv \mathrm{NFP}_3 + \mathcal{I}$). But $\mathrm{NFP}_3(\mathrm{pair})$ proves that $M = \{(\{x\}, y) \mid x \in y\}$ is a set and

$$\forall s \exists a \forall x(x \in a \leftrightarrow \exists q(q \in s \wedge (\forall z \in x)((q, z) \in M))). \tag{7}$$

Choose $s = \{\{x\} \mid x \in V\}$ in (7); then $\mathrm{NFP}_3(\mathrm{pair})$ proves that there is a set I such that

$$(\forall x)(x \in I \leftrightarrow \cap x \neq \emptyset). \qquad \square$$

It follows that we can freely use the homogeneous pairing operation when we work in NF and NFI.

[17] A similar argument shows that NFI justifies the existence of the largest fixed point of Γ_A.

4. Generating a Truth Set in NFI

We now simulate via homogeneous pairing the logical operations, which are essential for introducing in the Quinean universe a counterpart of the formula-representing map of §1.1.

Definition 7.

$$\begin{aligned} \dot{\neg}x &:= (0, x);\\ x\dot{\wedge}y &:= (1, (x, y));\\ \dot{\forall}f &:= (2, f);\\ \dot{\in}xy &:= (3, (x, y)). \end{aligned}$$

Then we can inductively introduce a map $A \mapsto [A]$ such that $FV(A) = FV([A])$ and

- $[t \in s] := \dot{\in}\{t\}s$;
- $[\neg A] := \dot{\neg}[A]$;
- $[A \wedge B] := \dot{\wedge}[A][B]$;
- $[\forall x\, A] := \dot{\forall}\{x \mid A\}$.

We also define

$$y \cdot x := [x \in y].$$

Under the dot-application, the universe of sets becomes an applicative structure. Note however that $y \cdot x$ is stratified only if y and x are given the types $i+1$ and i (respectively), and that the result of applying y to x is one greater than the type of x.

Lemma 11 (Restricted diagonalization, in NFI). *There is a term R such that*

$$R = \dot{\neg}R\,.$$

Moreover, for every a, there is a term $\Delta(a)$ such that

$$\Delta(a) = [\neg(a \in \Delta(a))].$$

Proof. The operation $\Gamma(a) = (\dot{\neg}, a)$ is monotone in a. By the fixed point lemma, there is some R satisfying the claim. As to the second part, $\Phi_a(b) = [a \in b] = (3, (a, b))$ is monotone in b. So the conclusion is implied by lemma 10. □

We now model the Kripke–Feferman notion of self-referential truth, as developed in [3], within the abstract framework of Quine's set theory. The truth predicate $\mathcal{T}$ is introduced as the fixed point of a stratified positive (in a) operator $\mathcal{T}(x, a)$, which

encodes the recursive clauses for partial self-referential truth and is given by the formula

$$\begin{aligned}\exists u\exists v\exists w\ [\ & (x = [u \in v] \wedge u \in v) \vee \\ & \vee\ (x = [\neg u \in v] \wedge \neg u \in v) \vee \\ & \vee\ (x = [\neg\neg v] \wedge\ v \in a) \vee \\ & \vee\ (x = v \dot{\wedge} w \wedge\ v \in a \wedge w \in a) \vee \\ & \vee\ (x = \dot{\neg}(v \dot{\wedge} w) \wedge\ (\dot{\neg} v \in a \vee\ \dot{\neg} w \in a)) \vee \\ & \vee\ (x = \dot{\forall} v \wedge \forall z(v \cdot z\ \in a)) \vee \\ & \vee\ (x = \dot{\neg}\dot{\forall} v \wedge \exists z(\dot{\neg} v \cdot z \in a))].\end{aligned}$$

Clearly $\{x \mid \mathcal{T}(x, a)\}$ is $\subseteq$-monotone in a and is predicative: it receives type 2 once we assign type 0 to u, z, type 1 to x, v, w, type 2 to a.

Definition 8.

$$\begin{aligned} Cl_T(a) &:= \forall x(\mathcal{T}(x, a) \to x \in a); \\ T &:= \{x \mid \forall a(Cl_T(a) \to x \in a)\}; \\ Tx &:= x \in T. \end{aligned}$$

Lemma 10 immediately implies:

Proposition 8. NFI *proves:*

1. $\exists y(y = T)$;
2. $\forall x(\mathcal{T}(x, T) \to x \in T)$;
3. $Cl_T(a) \to T \subseteq a$.

Proposition 9. NFI *proves:*

$$\begin{aligned} T[x \in y] &\leftrightarrow x \in y; \\ T[\neg x \in y] &\leftrightarrow \neg x \in y; \\ T\dot{\neg}\dot{\neg} x &\leftrightarrow Tx; \\ T(x \dot{\wedge} y) &\leftrightarrow Tx \wedge Ty; \\ T\dot{\neg}(x \dot{\wedge} y) &\leftrightarrow T(\dot{\neg} x) \vee T(\dot{\neg} y); \\ T\dot{\forall} f &\leftrightarrow \forall x T(f \cdot x); \\ T\dot{\neg}\dot{\forall} f &\leftrightarrow \exists x T\dot{\neg}(f \cdot x). \end{aligned}$$

Proof. Use $Tx \leftrightarrow \mathcal{T}(x, T)$, which is provable with proposition 8, and the independence properties of $\dot{\neg}$, $\dot{\wedge}$, $\dot{\forall}$, $\dot{\in}$, which follow from lemmata 8–9. □

Hence, as interesting special cases, we obtain:

$$(T[Tx] \leftrightarrow Tx) \ \wedge \ (T[\neg Tx] \leftrightarrow \neg Tx). \tag{8}$$

Proposition 10 (Consistency). NFI *proves:*

1. $\forall x(Tx \rightarrow \neg T\dot{\neg}x)$;
2. $\exists y(\neg Ty \wedge \neg T\dot{\neg}y)$.

Proof. Ad 1: choose $\psi(a) := \neg T(\dot{\neg}a)$; $\{x \mid \psi(x)\}$ exists in NFI. Then check:

$$\forall a(\mathcal{T}(a, \{x \mid \psi(x)\}) \ \rightarrow \ \psi(a)).$$

Ad 2: let $y = R$ (lemma 11) and apply consistency. □

Remark 3. Consider the map:

$$x \mapsto \varphi(x) := [\neg Tx].$$

An alternative "Liar propositional object" would be a set L such that

$$L = [\neg TL] = [\neg(L \in T)].$$

But *the above equation cannot be stratified*; indeed, by (8) NFI proves:

$$\neg\exists x(x = [\neg Tx]).$$

Lemma 12. *If A is stratified* (*and $\{x \mid A\}$ is loosely predicative*)*, then*

$$T[\forall xA] \ \leftrightarrow \ \forall xA;$$
$$T[\neg\forall xA] \ \leftrightarrow \ \neg\forall xA$$

are provable in NF (NFI).

Proof. Consider the following steps:

$$\begin{aligned} T[\forall xA] &\leftrightarrow \forall uT(\{x \mid A\} \cdot u); \\ &\leftrightarrow \forall uT[u \in \{x \mid A\}]; \\ &\leftrightarrow \forall u(u \in \{x \mid A\}); \\ &\leftrightarrow \forall uA[x := u]. \end{aligned}$$

Observe that the second step uses proposition 9, while the last step requires stratified comprehension in NF (or in NFI, provided A is loosely predicative). □

Theorem 8 (Stratified T-schema). *If A is stratified,* NF *proves:*

$$T[A] \leftrightarrow A.$$

If A is $\forall$-free, the schema is already provable in NFI.

Proof. By induction on A with proposition 9 and the previous lemma. □

The stratified T-schema implies that T strongly deviates from the behaviour of self-referential truth predicates à la Kripke–Feferman, which do not in general validate the truth axioms themselves nor *arbitrary* logical axioms, see [16] and [8]. On the contrary, *T provably believes that it is two-valued, consistent and that it satisfies the closure conditions embodied by the operator formula $\mathcal{T}(x, T)$ generating partial truth itself*; this marks the difference of the present theory versus the Kripke–Feferman system KF:

Corollary 2. NFI *proves:*

$$T[Ta \ \vee \ \neg Ta]; \tag{9}$$
$$T[\neg(Ta \ \wedge \ \neg Ta)]; \tag{10}$$
$$T[T[x \in y] \leftrightarrow x \in y]; \tag{11}$$
$$T[T[\neg x \in y] \leftrightarrow \neg x \in y]; \tag{12}$$
$$T[T\dot{\neg}\dot{\neg}x \leftrightarrow Tx]; \tag{13}$$
$$T[T(x\dot{\wedge}y) \leftrightarrow Tx \wedge Ty]; \tag{14}$$
$$T[T\dot{\neg}(x\dot{\wedge}y) \leftrightarrow T(\dot{\neg}x) \vee T(\dot{\neg}y)]. \tag{15}$$

In addition, NF proves:

$$T[\forall x T(f \cdot x) \leftrightarrow T\dot{\forall} f]; \tag{16}$$
$$T[\forall x(Tx \leftrightarrow \mathcal{T}(x, T))]. \tag{17}$$

Proof. All the claimed formulas within the scope of the outermost T are provable in NF (at most):

- (9)–(10) trivially by logic;
- (11)–(16) by proposition 9;
- (17) by proposition 8.

Also, the same formulas are stratified. Hence the conclusions of the corollary follow by the T-stratified schema (theorem 8). □

4.1. Final Remarks

Definition 9.

$$Pa :\Leftrightarrow Ta \vee T\dot{\neg}a\,.$$

Pa formally represents the predicate "a is a proposition".

Proposition 11 (NFI). *The collection of all propositions is a proper subset of the universe:*

$$\exists y(y = \{x \mid Px\}) \wedge \{x \mid Px\} \subset V.$$

Moreover P has the following closure properties:

$$\begin{aligned} P([x \in y]) \ &\wedge \ P[Tx]; \\ Pa \wedge (Ta \to Pb) \ &\to \ P(a \dot{\to} b); \\ Pa \wedge Pb \ &\to \ P(a \dot{\wedge} b) \wedge P(a \dot{\vee} b); \\ Pa \ &\to \ T[Pa]; \\ &\quad P(\dot{\forall} f). \end{aligned}$$

The first claim is a consequence of proposition 10. As for the remaining properties, apply proposition 9.

A few closure conditions of the previous proposition are reminiscent of corresponding axioms in PT and AT; but we stress that the axiom $T[\neg Px]$ of AT is refuted in the present context, while the (analogue of the) last statement is clearly unsound, provably in AT and PT. Note also that $P(\dot{\forall} f)$ implies $\forall x\, P(f \cdot x)$, i.e., *every set defines a propositional function*.

We conclude by representing Russell's contradiction within the theory of propositions and truth that we have so far developed in NFI.

Definition 10.

$$\tau(f) := [P(\dot{\forall} f)].$$

By definition of the map $A \mapsto [A]$, pairing, and proposition 11, we obtain:

Lemma 13. NFI *proves:*

$$\begin{aligned} P(\tau(f)) \ &\wedge \ T(\tau(f)); \\ \tau(f) = \tau(g) \ &\to \ f = g\,. \end{aligned}$$

So the operation τ is a well-defined injective map from sets into truths (and propositions).

Proposition 12. NFI *proves:*

$$\neg\, \exists d \forall x(x \in d \leftrightarrow (\exists f \subseteq P)(x \notin f \wedge x = \tau(f))).$$

To sum up: we have considered three formal systems for dealing with Russell's contradiction in Appendix B of [20]. In all systems the Russellian argument can be naturally formalized, but it is essentially sterilized either by denying the existence of a suitable propositional function or a set. The first two systems are provably non-extensional and predicatively inclined; no analogue of the power set is apparently

definable, while the collections of truths and propositions do not define completed totalities.

In the third system we do have the set of all propositions and truths, and extensionality is basic. The contradiction is then avoided by the mechanism of stratification and in accordance with an impredicative type theoretic perspective.

Acknowledgement. Research supported by Università degli Studi di Firenze and MIUR. Part of this paper has been presented under the title *Relating KF to NF* at the Symposium *Russell 2001*, June 2-5 2001, München.

References

[1] Aczel, Peter: 1980. Frege structures and the notions of proposition, truth and set. In: J. Barwise *et al*. (eds.), *The Kleene Symposium*, Amsterdam: North-Holland, 31–59.

[2] Barendregt, Henk: 1984. *The Lambda Calculus: its Syntax and Semantics*. Amsterdam: North-Holland.

[3] Cantini, Andrea: 1996. *Logical Frameworks for Truth and Abstraction*. Amsterdam: North-Holland.

[4] Cantini, Andrea and Pierluigi Minari: 1999. Uniform inseparability in explicit mathematics. *The Journal of Symbolic Logic* 64: 313–326.

[5] Cocchiarella, Nino: 2000. Russell's paradox of the totality of propositions. *Nordic Journal of Philosophical Logic* 5: 23–37.

[6] Crabbè, Marcel: 1982. On the consistency of an impredicative subsystem of Quine's NF. *The Journal of Symbolic Logic* 47: 131–136.

[7] Feferman, Solomon: 1978. Constructive theories of functions and classes. In: M. Boffa *et al*. (eds.), *Logic Colloquium '78*, Amsterdam: North-Holland, 55–98.

[8] Feferman, Solomon: 1991. Reflecting on incompleteness. *Journal of Symbolic Logic* 56: 1–49.

[9] Forster, Thomas: 1995. *Set Theory with a Universal Set*. Oxford Logic Guides, 31. Oxford: Clarendon.

[10] Glass, Thomas: 1996. On power class in explicit mathematics. *Journal of Symbolic Logic* 61: 468–489.

[11] Girard, Jean Y.: 1987. *Proof Theory and Logical Complexity*. Napoli: Bibliopolis.

[12] Grim, Peter: 1991. *The Incomplete Universe. Totality, Knowledge and Truth*. Cambridge, MA: MIT Press.

[13] Grishin, V. N.: 1969. Consistency of a fragment of Quine's NF system. *Soviet Mathematics Doklady* 10: 1387–1390.

[14] Holmes, Randall M.: 1995. The equivalence of NF-style set theories with "tangled" type theories: the construction of ω-models of predicative NF (and more). *The Journal of Symbolic Logic* 60: 178–190.

[15] Kalish, Donald and Richard Montague: 1964. *Logic. Techniques of Formal Reasoning*. New York: Harcourt, Brace and World.

[16] Kripke, Saul: 1975. Outline of a theory of truth. *Journal of Philosophy* 72: 690–716.

[17] Oksanen, Mika: 1999. The Russell-Kaplan paradox and other modal paradoxes; a new solution. *Nordic Journal of Philosophical Logic* 4: 73–93.

[18] Quine, Willard V.: 1945. On ordered pairs. *The Journal of Symbolic Logic* 10: 95–96.

[19] Rosser, John B.: 1953. *Logic for Mathematicians*. New York: McGraw-Hill.

[20] Russell, Bertrand: 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press. Reprinted by Routledge: London, 1997.

[21] Specker, Ernst: 1953. The axiom of choice in Quine's *New Foundations for Mathematical Logic*. *Proceedings of the National Academy of Sciences of the U.S.A.* 39: 972–975.

[22] Specker, Ernst: 162. Typical ambiguity. In: E. Nagel *et al.* (eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, Stanford: Stanford University Press, 116–123.

Dipartimento di Filosofia
Università degli Studi di Firenze
via Bolognese 52
50139 Firenze
Italy

E-mail: cantini@philos.unifi.it

The Consistency of the Naive Theory of Properties

Hartry Field

Abstract. The Naive Theory of Properties asserts, for any predicate $\Theta(x)$, that there is a property with the feature that an object o has this property if and only if $\Theta(o)$. If properties are to play a useful role in semantics it is hard to avoid assuming the Naive Theory, yet it appears to lead to various paradoxes. The paper shows that no paradoxes arise from the Naive Theory as long as the logic is weakened appropriately, e.g., by avoiding excluded middle and using an appropriate conditional; the main difficulty is finding a semantics that can handle a conditional obeying reasonable laws without engendering paradox. (The semantics that's employed is infinite-valued, with the values only partially ordered.) The paper also discusses whether the solution can be adapted to Naive Set Theory; the answer is 'probably not', but it is argued that limiting Naive Comprehension in set theory is perfectly satisfactory, whereas doing so in a property theory used for semantics is not.

1. Introduction

According to the Naive Theory of Properties, for every predicate $\Theta(x)$ there is a corresponding property $\lambda x \Theta(x)$. Moreover, this property $\lambda x \Theta(x)$ is instantiated by an object o if and only if $\Theta(o)$. More generally, the Naive Theory involves the following "Naive Comprehension Schema":

$$\forall u_1 \ldots \forall u_n \exists y [\text{Property}(y) \wedge \forall x (x \textit{ instantiates } y \leftrightarrow \Theta(x, u_1, \ldots, u_n))]. \qquad \textbf{(NC)}$$

This Naive Theory of Properties has many virtues, but it seems to have been shattered by (the property version of) Russell's Paradox.

"Seems to" have been shattered? There's no doubt that it *was* shattered, if we presuppose full classical logic. Let us use the symbol '$\in$' to mean "instantiates". The Russell Paradox involves the Russell property R corresponding to the predicate 'does not instantiate itself'. So according to the Naive Theory, $\forall x [x \in R \leftrightarrow \neg(x \in x)]$. Therefore in particular,

$$R \in R \leftrightarrow \neg(R \in R). \qquad (*)$$

But (*) is classically inconsistent.

There are two solution routes (routes for modifying the Naive Theory) within classical logic. The first says that for certain predicates, such as 'does not instantiate itself', there is no corresponding property. The second says that there is one, but it isn't instantiated by what you might think: there are either (i) cases where an object o has the property $\lambda x \Theta(x)$ even though $\neg\Theta(o)$, or (ii) cases where an object o doesn't have the property $\lambda x \Theta(x)$ even though $\Theta(o)$. In particular, when $\Theta(x)$ is 'does not instantiate itself', the Russell property is either of sort (i) or sort (ii). This second solution route subdivides into three variants. One variant commits itself to a solution of type (i): the Russell property instantiates itself, but nonetheless has the property of not instantiating itself. A second variant commits itself to a solution of type (ii): the Russell property doesn't instantiate itself, but nonetheless fails to have the property of not instantiating itself. A third variant hedges: it says that the Russell property is either of sort (i) or sort (ii), but refuses to say which.

These four classical theories—the three variants that admit the existence of the Russell property and the one that denies it—all seem to me problematic. (In the *prima facie* analogous case of sets, I take the approach that denies the existence of "the Russell set" to be quite *un*problematic. But I take properties to be very different from sets in this regard, for reasons to be discussed in the final section.) In my view we need a different sort of solution route, and it must inevitably involve a weakening of classical logic. It is the aim of this paper to provide one.

The idea of weakening logic to avoid the Russell Paradox is not new, but the proposal presented here is unlike many in that it saves the full Naive Comprehension Schema in the form stated above: it saves it not only from the Russell Paradox (which is relatively easy) but from far more virulent forms of paradox (such as the Curry Paradox and its many extensions). I know of no other ways of saving Naive Comprehension in as strong or as natural a logic.

2. Background

If we are going to weaken classical logic to get around the Russell paradox (along with others), it is useful to look at how it is that (*) leads to contradiction in classical logic; that way, we'll know which steps in the argument for contradiction might be denied. (Actually, one well-known approach accepts contradictions, in the sense of assertions of form $A \wedge \neg A$, and while I do not favor it, I want my initial discussion to recognize it as an alternative. For that reason, let me stipulate that a theory is to be called *inconsistent* if it implies, not just a contradiction in the above sense, but anything at all: the existence of Santa Claus, the omniscience of George Bush about matters of quantum field theory, you name it. So even those who accept "contradictions" won't want their theory to be "inconsistent", in the way I am now using these terms.) With this terminology in mind, here are the main steps in an obvious argument that (*) is inconsistent:

(1) (*) and $R \in R$ together imply the contradiction

$$(R \in R) \wedge \neg(R \in R), \qquad (**)$$

since the first conjunct is one of the premises and the second conjunct follows by modus ponens.

(2) Analogously, (*) and $\neg(R \in R)$ together imply (**).

(3) So by disjunction elimination, (*) and $(R \in R) \vee \neg(R \in R)$ together imply the contradiction (**).

(4) But $(R \in R) \vee \neg(R \in R)$ is a logical truth (law of excluded middle), so (*) *all by itself* implies the contradiction (**).

(5) Anything that implies a contradiction implies anything whatever, and hence is inconsistent in the most obviously odious sense of the term.

That's the argument. I've been a bit sloppy about use and mention, since I've defined R to be a property, but appear to have spoken of a sentence (*) that contains it. There are several ways this could be made right. One is to work in a language where we have a property-abstraction operator, so that we could name R in the language; then that name would be used in (*). A second is to replace the 'R' in (*) with a free variable y: then the argument in the text goes over to an argument that formulas of form $y \in y \leftrightarrow \neg(y \in y)$ imply contradictions, so their existential generalizations do too, and (NC) implies such an existential generalization. A third involves the introduction of a convention of "parameterized formulas": pairs of formulas and assignments of objects to their free variables. Then (*) is simply a convenient notation for the pair of '$y \in y \leftrightarrow \neg(y \in y)$' and an assignment of R to 'y', and what appears in the text is a literally correct derivation involving parameterized formulas. Do things however you like.

Obviously there are several different possible ways to restrict classical logic so as to evade the above argument for the inconsistency of (*). (I take it that the argument that (NC) implies (*) involves nothing in the least controversial, so that it is the argument that (*) leads to inconsistency that must be challenged.) I will simply state my preferred approach, without arguing that it is best: in my view, the most appealing way to weaken classical logic so as to evade the argument that (*) leads to inconsistency is to restrict the law of excluded middle, thereby undermining Step (4). Disjunction elimination can be retained (even in the strong sense used in Step (3), i.e., allowing side formulas). So can "the odiousness of contradictions" assumed in Step (5).

Of course, if you can evade the argument in a logic $\pounds_1$ that contains the "odiousness of contradictions" rule $A \wedge \neg A \models B$, you can equally evade it in a logic $\pounds_2$ just like $\pounds_1$ but which is "paraconsistent" in that the rule $A \wedge \neg A \models B$ is dropped. But since classical laws like excluded middle that are absent from $\pounds_1$ will be absent from $\pounds_2$ as well, this has no evident advantages. What might have advantages, if it could be achieved, would be to save Naive Property Theory in a paraconsistent logic in which we keep laws absent from $\pounds_1$, such as excluded middle. I know of no interesting way to achieve this; for a discussion of some obstacles, see [2].

Unfortunately, restricting excluded middle falls far short of giving an adequate theory. Restricting the law of excluded middle blocks *the above argument* for the inconsistency of $R \in R \leftrightarrow \neg(R \in R)$, but it is by no means obvious that there is a satisfactory logic without unrestricted excluded middle in which that biconditional can be maintained. It is still less obvious that there is a satisfactory such logic in which the full Naive Theory of Properties can be maintained. Let me explain.

The most obvious ways to deal with the paradoxes in logics without excluded middle (e.g., the property-theoretic adaptation of the Kleene version of Kripke's [4] "fixed point" approach to the semantic paradoxes)[1] do not vindicate (NC), nor do they even vindicate its weak consequence (*). The reason is that they don't contain an appropriate conditional (or biconditional).

Indeed, the main issues involved in showing the consistency of the Naive Theory center on the problem of finding an adequate treatment of the $\rightarrow$. (And hence the $\leftrightarrow$: I'll assume that $A \leftrightarrow B$ means $(A \rightarrow B) \wedge (B \rightarrow A)$.) Even if our goal were limited to the consistent assertion of the biconditional (*), that would pretty much rule out our defining $A \rightarrow B$ in terms of the other connectives in the manner familiar from classical logic, viz., $\neg A \vee B$. For on that "material conditional" reading of '$\rightarrow$', (*) amounts to

$$[\neg(R \in R) \vee \neg(R \in R)] \wedge [\neg\neg(R \in R) \vee (R \in R)].$$

Assuming distributivity and a few other simple laws, this is equivalent to a disjunction of the classical inconsistencies $(R \in R) \wedge \neg(R \in R)$ and $\neg(R \in R) \wedge \neg\neg(R \in R)$. If we assume double-negation elimination, that's in effect just the simple contradiction $(R \in R) \wedge \neg(R \in R)$; and even if double-negation elimination isn't assumed, a disjunction of contradictions seems just as inconsistent as a single contradiction. So if we put aside the paraconsistent approaches mentioned above, it's clear that we cannot in general interpret $A \rightarrow B$ as $\neg A \vee B$ if we want to retain even (*). And on the paraconsistent approaches the "material conditional" reading of $\rightarrow$ seems inappropriate on different grounds: that reading invalidates modus ponens. (Although it is important not to interpret $\rightarrow$ as the material conditional, the theory that I will advocate does posit a close relation between the two: while $(A \rightarrow B) \leftrightarrow (\neg A \vee B)$ is not a logical truth, it is a logical consequence of the premises $A \vee \neg A$ and $B \vee \neg B$. In other words, it is only in the context of a breakdown in the law of excluded middle that the divergence between the $\rightarrow$ and the material conditional emerges.)

The first problem about getting a decent conditional, then, is licensing the assertion of (*). But there are plenty of "logics of $\rightarrow$" that solve that problem while still being inadequate to the Naive Theory, for the full Comprehension Schema (NC) is not consistently assertable in them. Indeed, many of these logics fail to handle a close analog of Russell's paradox due to Curry. The problem is this: (NC) implies the

[1] Such an adaptation of the Kleene variant of Kripke's approach is in effect given in Maddy [5], as a theory of proper classes. I will discuss the use of the theory to be given in this paper in connection with proper classes in the final section. In my opinion, the presence of (NC) is not only needed for property theory generally, it also makes for a more adequate theory of proper classes.

existence of a Curry property K, for which $\forall x[x \in K \leftrightarrow (x \in x \rightarrow \bot)]$, where $\bot$ is any absurdity you like. So $K \in K \leftrightarrow (K \in K \rightarrow \bot)$; that is,

(i) $K \in K \rightarrow (K \in K \rightarrow \bot)$

and

(ii) $(K \in K \rightarrow \bot) \rightarrow K \in K$.

But in many logics of $\rightarrow$ we have the contraction rule $A \rightarrow (A \rightarrow B) \models A \rightarrow B$, on which (i) implies

(i*) $K \in K \rightarrow \bot$.

But this with (ii) leads to $K \in K$ by modus ponens; and another application of modus ponens leads from that and (i*) to $\bot$.

Unless we restrict modus ponens (and it turns out a very drastic restriction of it would be required), we need to restrict the contraction rule. This requires further restrictions on the logic as well. For instance, given that we're keeping modus ponens in the form $A, A \rightarrow B \models B$, we certainly have $A, A \rightarrow (A \rightarrow B) \models B$ simply by using modus ponens twice; so to prevent contraction, we certainly can't have the generalized $\rightarrow$-introduction meta-rule that allows passage from $\Gamma, A \models B$ to $\Gamma \models A \rightarrow B$. Indeed, even the weaker version which allows the inference only when Γ is empty should be given up: it is the obvious culprit in an alternative derivation of the Curry paradox.

It turns out, though, that the difficulty in finding an adequate treatment of the '$\rightarrow$' is not insuperable, and that the Naive Comprehension Principle (NC) can be maintained; indeed, it can be maintained in a logic that, though not containing excluded middle or the contraction rule, is not altogether unnatural or hopelessly weak.[2] The aim of this paper is to show this.[3] Whether the theory should still count as "naive" when the logic is altered in this way is a question I leave to the reader.

It is worth emphasizing that though the law of excluded middle will need restriction, there is no need to give it up entirely: it can be retained in various restricted circumstances. For instance, the notion of property is normally employed in connection with a "base language" L that does not talk of properties; we then expand L to a language L^+ that allows for properties, including but not limited to properties of things talked about in L. (It is not limited to properties of things talked about in L because it will also include properties of properties: indeed, it is some of these that give rise to the apparent paradoxes.) It is within the ground language L that most of mathematics, physics, and so forth takes place; and the theory advocated here does

[2]For instance, the conditional obeys contraposition in the strong form $\models (\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$. Also when $\models A \leftrightarrow B$ and C_B results from C_A by substituting B for one or more occurrences of A, then $\models C_A \leftrightarrow C_B$; so (NC) yields that $y \in \lambda x \Theta(x)$ is everywhere intersubstitutable with $\Theta(y)$, even within the scope of a conditional.

[3]The approach I'll be giving is an adaptation of the approach to the semantic paradoxes developed in [1].

not require any limitation of excluded middle in these domains, because as long as we restrict our quantifiers to the domain of the ground language we can retain full classical logic. We can also retain full classical logic in connection with those special ("rank 1") properties that are explicitly limited so as to apply to non-properties; and to those special ("rank 2") properties that are explicitly limited so as to apply to non-properties and rank 1 properties; and so on. Where excluded middle cannot be assumed is only in connection with certain properties that do not appear anywhere in such a rank hierarchy, like the Russell Property and the Curry Property (though for other such properties, e.g., those whose *complement* appears in the rank hierarchy, excluded middle is also unproblematic). Even for the "problematic" properties, there is no need to give up excluded middle for claims about property *identity*; it is only when it comes to claims about *instantiation* of problematic properties that excluded middle will not be able to be assumed in general.

I don't know if the theory here can be adapted to a theory of "naive sets", by adding an axiom or rule of extensionality; I will have a bit to say about this in the final section, including a discussion of why the matter is much less pressing for sets than for properties. But if it is possible to develop a theory of naive sets, it seems unlikely that we would be able to maintain excluded middle for identities between naive sets (e.g., between the empty set and $\{x|x = x \wedge K \in K\}$, where K is the "Curry set", defined in analogy with the Curry property). Because of this, a "naive set theory", if possible at all, would have an importantly different character from the naive property theory about to be developed.

3. The Goal

I've said I want a consistent naive theory of properties, but actually what I want is a bit stronger than mere consistency. It's time to start being a bit more precise.

Let L be any first order language with identity. Since I won't want to identify $A \rightarrow B$ with $\neg A \vee B$, it is necessary to assume that $\rightarrow$ is a primitive connective, along with $\neg$, $\wedge$ and/or $\vee$, and $\forall$ and/or $\exists$. And to avoid annoying complications about how to extend function symbols when we add to the ontology, I'll assume that L contains no function symbols (except perhaps for 0-place ones, i.e., individual constants). L can be taken to be a language for mathematics, or physics, or whatever you like other than properties. (So it shouldn't contain the terms 'Property' or '$\in$' in the senses to be introduced. If it contains these terms in other senses—e.g., '$\in$' for membership among the iterative sets of standard set theory—then imagine these replaced by other terms.)

Let L^+ result from L by adding a new 1-place predicate 'Property' and a new 2-place predicate '$\in$' meaning 'instantiates'. For any formula A of L, let A^L be the formula of L^+ obtained from A by restricting all bound occurrences of any variable z by the condition '$\neg$Property(z)'. Let T be any theory in the language L. "Naive

Property Theory over T" is the theory T^+ that consists of the following non-logical axioms:

(I) A^L, for any A that is a closure of a formula that follows from T

(II) $\forall x \forall y[x \in y \rightarrow \text{Property}(y)]$

(III) $\forall u_1 \ldots \forall u_n \exists y[\text{Property}(y) \wedge \forall x(x \in y \leftrightarrow \Theta(x, u_1, \ldots, u_n))]$,
where $\Theta(x, u_1, \ldots, u_n)$ is any formula of L^+ in which y is not free.

(III) is just (NC). Then a minimal goal is to show that in a suitable logic, the theory T^+ consisting of (I)–(III) is always consistent as long as T itself is consistent. Note that if T is itself a classical theory, i.e., is closed under classical consequence, then "Naive Property Theory over T" effectively keeps classical logic among sentences of the form A^L, even though its official logic is weaker: for if A_1,...,A_n are formulas of L that classically entail B, then $A_1 \wedge A_2 \wedge \ldots A_n \rightarrow B$ is in T, so $[(\forall u_1, \ldots, u_k)(A_1 \wedge A_2 \wedge \ldots A_n \rightarrow B)]^L$ is in T^+, and this is the same as $(\forall u_1, \ldots, u_k)[\neg\text{Property}(u_1) \wedge \ldots \wedge \neg\text{Property}(u_k) \rightarrow (A_1^L \wedge A_2^L \wedge \ldots A_n^L \rightarrow B^L)]$.

The *minimal* goal is to show that T^+ is consistent whenever T is, but I actually want something slightly stronger: I want to introduce a kind of multi-valued model for L^+ (infinite-valued, in fact), and then prove

(G) *For each classical model M of L, there is at least one model M^+ of L^+ that validates* (II) *and* (III) *and has M as its reduct;*

where to say that M is the reduct of M^+ means roughly that when you restrict the domain of M^+ to the things that don't satisfy 'Property' (and forget about the assignments to 'Property' and to '$\in$') then what you are left with is just M.[4] Since the connectives of L^+ will reduce to their classical counterparts on the reduct, the fact that M is the reduct of M^+ will guarantee the validity of Axiom Schema (I); so if M satisfies T, M^+ satisfies T^+.

There is good reason why (G) says 'at least one' rather than 'exactly one': we should expect that most or all models of T^+ can be extended to models that contain new properties but leave the property-less reduct unchanged. The proof that I will give yields the minimal M^+ for a given M, but extensions of the model with the same reduct could easily be given.

I will prove (G) in a classical set-theoretic metalanguage, so anyone who is willing to accept classical set-theory should be able to accept the coherence of the non-classical property theory to be introduced.

[4] The reason for the 'roughly' in the definition of 'reduct' is that M is a classical model, whereas M^+ will be multi-valued; so its reduct will have to assign objects that live in the larger space of values. Nonetheless, the larger space of values will contain two rather special ones, to be denoted **1** and **0**, and we can take 'A has value **1** in M^+' and 'A has value **0** in M^+' to correspond to 'A is true in M' and 'A is false in M', when A is in L. The reduct of M^+ won't strictly be M, but it will be the $\{\mathbf{0}, \mathbf{1}\}$-valued model that corresponds to M in the obvious way.

4. The Semantic Framework

The goal just enunciated calls for developing a model-theoretic semantics for L^+ in a classical set-theoretic metalanguage. The semantics will be multi-valued: in addition to (analogs of) the usual two truth values there will be others, infinitely many in fact.

4.1. The Space of Values

My approach to achieving the goal is an extension of the Kripke-style approach previously mentioned, but it needs to be substantially more complicated because of the need for a reasonable conditional.

One complication has to do with method of proof: the new conditional is not "monotonic" in the sense of Kripke, which means that we cannot make do merely with the sort of fixed point argument that is central to his approach (though such a fixed point argument will play an important role in this approach too).

The other complication is that the semantic framework itself should be generalized: whereas Kripke uses a 3-valued semantics, I will use a model theory in which sentences take on values in a subspace W^Π of the set F^Π of functions from $\mathrm{Pred}(\Pi)$ to $\{0, \frac{1}{2}, 1\}$, where Π is an initial ordinal (ordinal with no predecessor of the same cardinality) that is greater than ω, and where $\mathrm{Pred}(\Pi)$ is the set of its predecessors.[5] (I don't fix on a particular value of Π at this point, because I will later impose further minimum size requirements on it.)

Which subset of F^Π do I choose as my W^Π? If ρ is a non-zero ordinal less than Π, call a member f of F^Π *ρ-cyclic* if for all β and σ for which $\rho \cdot \beta + \sigma < \Pi$, $f(\rho \cdot \beta + \sigma) = f(\sigma)$; and call it *cyclic* if there is a non-zero ρ less than Π such that it is ρ-cyclic. Call it *regular* if in addition to being cyclic, it satisfies the condition that it is either one of the constant functions $\mathbf{0}$ and $\mathbf{1}$ (which map everything into 0 or map everything into 1) or else maps 0 into $\frac{1}{2}$. Then W^Π consists of the regular functions from $\mathrm{Pred}(\Pi)$ to $\{0, \frac{1}{2}, 1\}$. (Once we've found a suitable method of assigning values in W^Π to sentences, then the valid inferences among sentences will be taken to be those inferences that are guaranteed to preserve the value $\mathbf{1}$.)

A few properties of W^Π are worth noting.

- It has a natural partial ordering: $f \preceq g$ iff $(\forall \alpha < \Pi)(f(\alpha) \leq g(\alpha))$. The ordering has a minimum $\mathbf{0}$ and maximum $\mathbf{1}$. And the ordering is not total: for instance, the constant function $\mathbf{\frac{1}{2}}$ is incomparable with the function that has value $\frac{1}{2}$ at limit ordinals, 0 at odd ordinals, and 1 at even successors.

[5] When I presented an analogous solution to the semantic paradoxes in [1], I did not explicitly introduce this new space of semantic values (since I hadn't yet thought of the matter in that way); but the ideas seem to me clearer with this space of values made explicit.

- For each $f \in W^{\Pi}$ define $f^{\bigstar}$ to be the function for which $f^{\bigstar}(\alpha) = 1 - f(\alpha)$ for each α. Then $f^{\bigstar}$ will be in W^{Π} too. Moreover, the operation $\bigstar$ is a symmetry that switches $\mathbf{0}$ with $\mathbf{1}$, leaving the constant function $\frac{\mathbf{1}}{\mathbf{2}}$ fixed.

- For any nonempty subset S of W^{Π} that has cardinality less than that of Π, define $\curlywedge(S)$ to be the function whose value at each α is the minimum of $\{f(\alpha) | f \in S\}$, and $\curlyvee(S)$ to be the function that analogously gives the pointwise maximum. Then $\curlywedge(S)$ and $\curlyvee(S)$ are in W^{Π};[6] and clearly, they are the meet and join of S with respect to the partial ordering.

- For any S, $\curlyvee(S)$ is $\mathbf{1}$ only if $\mathbf{1} \in S$; that holds because if $\mathbf{1} \notin S$, then $f(0) < 1$ for each f in S. This is important: it will ensure that the logic that results will obey the meta-rules of $\vee$-elimination and $\exists$-elimination.

Observe also that if $f(0)$ is $\frac{1}{2}$ and $f \in W^{\Pi}$, then f assumes the value $\frac{1}{2}$ arbitrarily late, viz., at all right-multiples of ρ_f. (By ρ_f, I mean the smallest ρ for which f is ρ-cyclic.) Also, note that for any f and g in W^{Π}, there are $\rho < \Pi$ such that *both* f and g are ρ-cyclic: any common right-multiple of ρ_f and ρ_g will be one. One consequence of this is that if there are $\beta < \Pi$ for which $f(\beta) < g(\beta)$ (alternatively, $f(\beta) \leq g(\beta)$), then for any $\alpha < \Pi$, there are β in the open interval from α to Π for which $f(\beta) < g(\beta)$ (alternatively, $f(\beta) \leq g(\beta)$). And that implies that

- $f \preceq g$ is equivalent to the *prima facie* weaker claim that there is an α (less than Π) such that for all β greater than α (and less than Π), $f(\beta) \leq g(\beta)$.

Similarly, if we define a (quite strong) strict partial ordering $\prec\prec$ by $f \prec\prec g$ iff either ($f = \mathbf{0}$ and $g \succeq \frac{\mathbf{1}}{\mathbf{2}}$) or ($f \preceq \frac{\mathbf{1}}{\mathbf{2}}$ and $g = \mathbf{1}$), then

- $f \prec\prec g$ is equivalent to the *prima facie* weaker claim that there is an α (less than Π) such that for all β greater than α (and less than Π), $f(\beta) < g(\beta)$.

For if the "weaker" claim holds, then pick β to be a common right-multiple of ρ_f and ρ_g greater than α; since $f(\beta) < g(\beta)$, at least one of $f(\beta)$ and $g(\beta)$ isn't $\frac{1}{2}$, so at least one of $f(0)$ and $g(0)$ isn't $\frac{1}{2}$, so at least one of f and g is in $\{\mathbf{0}, \mathbf{1}\}$; and the rest is obvious.

[6] For $\curlywedge(S)$, this is trivial if one of the members of S is $\mathbf{0}$ or if S is $\{\mathbf{1}\}$. Otherwise, the value of $\curlywedge(S)$ at 0 is $\frac{1}{2}$, and we need only verify cyclicity. Let ρ_S be the smallest ordinal that is a right-multiple of all the ρ_f for $f \in S$; by the cardinality restriction on S, this is less than Π (and at least 2). Moreover, all members of S ρ_S-cycle, so $\curlywedge(S)$ ρ_S-cycles (and so $\rho_{\curlywedge(S)} \leq \rho_S$).

The results just sketched are the keys to proving a final feature of the space W^{Π}, that it is closed under the following operation $\Longrightarrow$:

If $\alpha > 0$, $(f \Longrightarrow g)(\alpha)$ is

1 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) \leq g(\gamma)$;

0 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) > g(\gamma)$;

$\frac{1}{2}$ otherwise,

and $(f \Longrightarrow g)(0)$ is

1 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $f(\gamma) \leq g(\gamma)$;

0 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $f(\gamma) > g(\gamma)$;

$\frac{1}{2}$ otherwise.

(The value $\frac{1}{2}$ can occur only at 0 and at limits.) Note that the exceptional treatment of 0 in effect turns the domain of the functions in W^{Π} into a "transfinite circle", in which 0 is identified with Π. And we clearly have

- $f \Longrightarrow g$ is **1** if and only if $f \preceq g$; and $f \Longrightarrow g$ is **0** if and only if $f \succ\succ g$.

Why is W^{Π} closed under $\Longrightarrow$? Since **1** and **0** are in W^{Π}, we need only show that when neither $f \preceq g$ nor $f \succ\succ g$ then $f \Longrightarrow g$ is regular. Let ρ be the smallest non-zero ordinal for which both f and g ρ-cycle, and let ρ^* be $\rho \cdot \omega$. I claim that $f \Longrightarrow g$ is ρ^*-cyclic, that is, for any $\sigma < \rho^*$, the value of $(f \Longrightarrow g)(\rho^* \cdot \delta + \sigma)$ is independent of δ; and that when σ is 0, $(f \Longrightarrow g)(\rho^* \cdot \delta + \sigma)$ is $\frac{1}{2}$. Case 1: $\sigma > 0$. Then $(f \Longrightarrow g)(\rho^* \cdot \delta + \sigma) = 1$ iff $(\exists \beta < \rho^* \cdot \delta + \sigma)(\forall \gamma)(\beta \leq \gamma < \rho^* \cdot \delta + \sigma \supset f(\gamma) \leq g(\gamma))$ iff $(\exists \beta)(\rho^* \cdot \delta \leq \beta < \rho^* \cdot \delta + \sigma)(\forall \gamma)(\beta \leq \gamma < \rho^* \cdot \delta + \sigma \supset f(\gamma) \leq g(\gamma))$; but that's independent of δ since f and g are (ρ-cyclic and hence) ρ^*-cyclic. Similarly for the δ-independence of the condition for $(f \Longrightarrow g)(\rho^* \cdot \delta + \sigma) = 0$. Case 2: $\sigma = 0$. We need that $(f \Longrightarrow g)(\rho^* \cdot \delta) = \frac{1}{2}$ for all δ. The reason is that for any $\alpha < \rho^* \cdot \delta$ (i.e., $\alpha < \rho \cdot (\omega \cdot \delta)$), there is an ζ such that $\alpha < \rho \cdot \zeta < \rho \cdot (\zeta + 1) < \rho \cdot (\omega \cdot \delta)$; and (since neither $f \preceq g$ nor $f \succ\succ g$) there are bound to be β in the interval from $\rho \cdot \zeta$ to $\rho \cdot (\zeta + 1)$ (lower bound included) where $f(\beta) > g(\beta)$ and others where $f(\beta) \leq g(\beta)$, so $(f \Longrightarrow g)(\rho^* \cdot \delta) = \frac{1}{2}$.

The operation $\Longrightarrow$ just specified has some rather nice properties. I've already noted the conditions under which it takes the values **1** and **0**. In addition:

- When f and g are in $\{\mathbf{0}, \mathbf{1}\}$ then $f \Longrightarrow g$ is identical to the value of the material conditional, $\Upsilon\{f^{\bigstar}, g\}$.

Since I will be using $\Longrightarrow$ to evaluate the conditional, this will mean that the conditional reduces to the material conditional when excluded middle is assumed for antecedent and consequent.

It is beyond the present scope to investigate the laws governing $\Longrightarrow$ (though this is important since it will determine which inferences involving $\rightarrow$ are valid); for that, see [1].

It's worth making explicit that if $f \Longleftrightarrow g$ is defined in the obvious way (as $\curlywedge\{f \Longrightarrow g, g \Longrightarrow f\}$), then if $\alpha > 0$, $(f \Longleftrightarrow g)(\alpha)$ is

1 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) = g(\gamma)$;
0 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $f(\gamma) \neq g(\gamma)$;
$\frac{1}{2}$ otherwise.

(And analogously for $(f \Longleftrightarrow g)(0)$: use Π in place of α on the right hand side.)

4.2. W^{Π}-Models

Having noted these features of the space W^{Π} of values, we can easily define models based on this space: W^{Π}-models. I will take a W^{Π}-*model* for a language to consist of a domain D *of cardinality less than* Π, an assignment to each individual constant c of a member $den(c)$ of D, and an assignment to each n-place predicate of a function p^* from D^n to W^{Π} (where D^n is the set of n-tuples of members of D). A W^{Π}-*valuation* for a language will consist of a W^{Π}-model together with a function s assigning objects in the domain of the model to the variables of the language. Given any valuation with assignment function s and any term t (individual constant or variable), let $den_s(t)$ be $den(t)$ if t is an individual constant, $s(t)$ if t is a variable.

Given a W^{Π}-valuation with assignment function s, we assign values in W^{Π} to formulas as follows:

$\|p(t_1, \ldots, t_n)\|_s$ is $p^*(den_s(t_1), \ldots, den_s(t_n))$, which in the future I'll also write as $p^*_{den_s(t_1),\ldots,den_s(t_n)}$;
$\|\neg A\|_s$ is $(\|A\|_s)^{\bigstar}$;
$\|A \wedge B\|_s$ is $\curlywedge\{\|A\|_s, \|B\|_s\}$;
$\|A \vee B\|_s$ is $\curlyvee\{\|A\|_s, \|B\|_s\}$;
$\|\forall x A\|_s$ is $\curlywedge\{\|A\|_{s'} \mid s'$ differs from s except perhaps in what is assigned to the variable $x\}$;
$\|\exists x A\|_s$ is $\curlyvee\{\|A\|_{s'} \mid s'$ differs from s except perhaps in what is assigned to the variable $x\}$;
$\|A \rightarrow B\|_s$ is $\Longrightarrow (\|A\|_s, \|B\|_s)$.

Note that for the quantifier clause to make sense in general, it is essential that the domain of quantification have lower cardinality than Π. But this is no real restriction, it's simply that if you want to consider models of large cardinality you have to choose a

large value of Π. (Recall the goal, (G): we want a strong form of consistency in which for any classical starting model M for the base language L, there is a non-classical model M^+ in L^+ that has M as its reduct. There is no reason why the space of values used for M^+ can't depend on the cardinality of M.) So I will take my non-classical model M^+ to be a W^{Π}-model for some initial ordinal Π of cardinality greater than that of M (as well as being greater than ω). M^+ will have a cardinality that is the maximum of the cardinalities of M and of ω, so this restriction will suffice for the quantifier clause to be well-defined.

I've written the valuation rules for ordinary formulas, but in the future I will adopt the convention of using parameterized formulas in which we combine the effect of the formula and the assignment function in our notation by plugging a metalinguistic name for an object assigned to a variable in for free occurrences of the variable in the displayed formula; that will allow me to drop the subscript s, and simplify the appearance of other clauses. For instance, the clauses for atomic formulas and universal quantifications become

$\|p(o_1, \ldots, o_n)\|$ is the function $f_{p,o_1,\ldots,o_n}$ that takes any α into
$p^*(o_1, \ldots, o_n)$;
$\|\forall x A\|$ is the function $\curlywedge\{\|A(o)\| \mid o \in D\}$.

(Sometimes I'll make the parameters explicit, e.g.,

For all $o_1, \ldots, o_n$, $\|\forall x A(o_1, \ldots, o_n)\|$ is the function
$\curlywedge\{\|A(o, o_1, \ldots, o_n)\| \mid o \in D\}$,

but the absence of explicit parameters should not be taken to imply that there are no parameters in the formula.)

5. A Model for Naive Property Theory

5.1. The Basics

The next step is to specify the particular model to be used for naive property theory. Recall that I'm imagining that we are given a model M for the base language L. We can assume without real loss of generality that $|M|$ (the domain of M) doesn't contain formulas of L^+, or n-tuples that include such formulas; for if the domain does contain such things, we can replace it with an isomorphic copy that doesn't. With this done, let E_0 be $|M|$. For each natural number k, we define a set E_{k+1} of *ersatz properties of level* $k+1$. A member of E_{k+1} is a triple consisting of a formula of L^+, a distinguished variable of L^+, and a function that assigns a member of $\bigcup\{E_j \mid j \leq k\}$ to each free variable of L^+ other than the distinguished one, meeting the condition that if $k > 0$ then at least one element of E_k is assigned.[7] If $\Theta(x, u_1, \ldots, u_n)$ is the formula

and x the distinguished variable and $o_1, \ldots, o_n$ the objects assigned to $u_1, \ldots, u_n$ respectively, I'll use the notation $\lambda x \Theta(x, o_1, \ldots, o_n)$ for the ersatz property. Let E be the union of all the E_k for $k \geq 1$, and let $|M^+|$ be $|M| \cup E$. (The cardinality of $|M^+|$ is thus the same as that of $|M|$, when $|M|$ is infinite, and is $\aleph_0$ when M is finite.) The only terms of L^+ besides variables are the individual constants of L; they get the same values in M^+ as in M.

I hope it's clear that the fact that I'm taking the items in the domain to be constructed out of linguistic items does not commit me to viewing properties as linguistic constructions; the point of the model is simply to give a strong form of consistency proof (i.e., to satisfy Goal (G)), and this is the most convenient way to do it.

Putting aside the unimportant issue of the nature of the entities in the domain, the domain does have a very special feature: all the properties in the model are ultimately generated (in an obvious sense I won't bother to make precise) from the entities in the ground model by the vocabulary of the ground model; so that the model contains the minimal number of properties that are possible, given the ground model. It is useful to consider such a special model for doing the consistency proof for naive property theory, but not all models of naive property theory will have this form (as is obvious simply from the fact that if we were to add new predicates to the ground model before starting the construction, we would generate new properties).

To complete the specification of M^+ we must specify an appropriate Π, and then assign to each n-place atomic predicate p a "W^Π-extension": a function p^* that takes n-tuples of members of $|M^+|$ into W^Π. I have already said that I would take Π to be the initial ordinal for a cardinal greater than the cardinality of $|M^+|$; a further stipulation will become necessary, but let us wait on that. As for the predicates, much of what we must say is obvious. If p is a predicate in L other than '=', and $o_1, \ldots, o_n$ are in M^+, we let $p^*_{o_1,\ldots,o_n}$ be $\mathbf{1}$ if $\langle o_1, \ldots, o_n \rangle$ is in the M-extension of p, $\mathbf{0}$ otherwise. (So it's $\mathbf{0}$ if any of the o_i are in E.) We let Property^*_o be $\mathbf{1}$ when o is in E, $\mathbf{0}$ otherwise. And we let $=_{o_1,o_2}$ be $\mathbf{1}$ when o_1 is the same object as o_2, and $\mathbf{0}$ otherwise. These stipulations obviously suffice to make M the reduct of M^+. Because of this, and the fact that the function assigned to each connective *including the conditional* reduces to its classical counterpart when confined to the set $\{\mathbf{0}, \mathbf{1}\}$, we get (by an obvious induction on complexity) that for any sentence A of L (or any formula A of L and any assignment function s that assigns only objects in $|M|$), the value of A^L in M^+ (relative to s) will be $\mathbf{1}$ when the value of A (relative to s) in M is 1, and will be $\mathbf{0}$ when the value of A (relative to s) in M is 0. Each instance of Axiom Schema (I) therefore gets value $\mathbf{1}$.

This leaves only '$\in$'. One desideratum should obviously be that when o_2 is in the ground model $|M|$, then $\in^*_{o_1,o_2}$ is $\mathbf{0}$. This will suffice for giving value $\mathbf{1}$ to Axiom (II).

The difficult matter, of course, is figuring out how to complete the specification of the W^Π-extension of '$\in$', in such a way as to validate Axiom Schema (III). This will be the subject of the next four subsections.

[7] The exception for $k = 0$ is needed only for formulas that contain no free variables beyond the distinguished one.

5.2. The Difficulty: How Do '$\in$' and '$\rightarrow$' Interact?

The main problem in constructing an interpretation for membership statements is due to the presence in the language of the conditional $\rightarrow$. Just to get a feeling for what might be involved here, consider a very simple case: the ordinary Curry property K. This is $\lambda x(x \in x \rightarrow \perp)$, where $\perp$ is some sentence with value $\mathbf{0}$, say $\exists y(y \neq y)$. What function $f_{K\in K}$ should serve as $\in^*_{K,K}$, and hence as $\|K \in K\|$, i.e., $\|K \in \lambda x(x \in x \rightarrow \perp)\|$? Since we want (III) to be valid, that had better be the same as $\|K \in K \rightarrow \perp\|$. That is, we want the function $f_{K\in K}$ to be identical to the function $f_{K\in K} \Longrightarrow \mathbf{0}$.

But how do we get that to be the case? The first thing we want to know is, what is $f_{K\in K}(0)$? The rules tell us that it is

$$1 \text{ if } (\exists\beta < \Pi)(\forall\gamma)[\beta \leq \gamma < \Pi \supset f_{K\in K}(\gamma) = 0];$$
$$0 \text{ if } (\exists\beta < \Pi)(\forall\gamma)[\beta \leq \gamma < \Pi \supset f_{K\in K}(\gamma) > 0];$$
$$\tfrac{1}{2} \text{ otherwise.}$$

It appears that we can't know the value of $f_{K\in K}(0)$ until we know the values of $f_{K\in K}(\alpha)$ for higher α. But finding out the values of $f_{K\in K}(\alpha)$ for each higher α seems to require already having the values for lower α. We seem to be involved in a vicious circle.

In fact, there is an easy way to find out what function $f_{K\in K}$ is. First, $f_{K\in K}(0)$ can't be 1; for the only function in W^Π that has value 1 at 0 is $\mathbf{1}$, and so $f_{K\in K}(1)$ would have to be 1; but $f_{K\in K}(1)$ can only be 1 if $f_{K\in K}(0)$ is 0. By a similar argument, $f_{K\in K}(0)$ can't be 0. It follows that $f_{K\in K}(0)$ must be $\frac{1}{2}$, and from that it is easy to successively obtain all the other values. (For the record, the value is $\frac{1}{2}$ at 0 and all limit ordinals, 0 at odd ordinals, and 1 at even successors.)

Other cases will not be so simple. For instance, consider a more general class of Curry-like properties, the properties of the form $\lambda x(x \in x \rightarrow A(x; o_1, \ldots, o_n))$. Letting Q be the property for a specific choice of A and of $o_1, \ldots, o_n$, we want $|Q \in Q|$ to have the same value as $Q \in Q \rightarrow A(Q; o_1, \ldots, o_n)$. But A can be a formula of arbitrary complexity, itself containing $\in$ and $\rightarrow$, and the o_is can themselves be "odd" properties of various sorts. It isn't obvious how the reasoning that works for the simple Curry sentence will work more generally.

In many specific cases, actually, it is also easy to come up with a consistent value for the sentences involved; often a unique one, though in cases like the parameterized sentence '$\lambda x(x \in x) \in \lambda x(x \in x)$' it is far from unique unless further constraints are added. But it's one thing to figure out what the value would have to be in a lot of individual cases, another to come up with a general proof that values always can be consistently assigned. And it's still another thing to specify a method that determines a unique value for any formula relative to any assignment function. How are we to do these further things? The reasoning about the valuation of $K \in K$ suggests that for $\alpha > 0$ we might be able to figure out the function Z_α that assigns to each parameterized formula B the value $\|B\|(\alpha)$, if only we had the functions Z_β

for $\beta < \alpha$; but getting the process going requires that we know Z_0, which includes the assignment to parameterized B for which there are arbitrarily high embeddings of '$\rightarrow$' in either the formula itself or the formulas involved in generating the parameters; this will depend on the Z_β for the very high values of β. The main problem, then, is to somehow break into the "transfinite circle".

I propose that we proceed by successive approximations. The main idea is to start *outside of* the space W^Π, so that we can treat '$\rightarrow$' in a way that mimics the behavior of the $\Longrightarrow$ at ordinal values greater than 0 but abandons its rigid requirement about stage 0. We will start out by assigning the "0^{th} stage" of the evaluation of all conditionals artificially, and see what the later stages must be like as a result of this; it will turn out that by continuing far enough we will inevitably be led to an appropriate assignment Z_0 of values for the initial stage.

This must be combined with another idea, which is basically the one Kripke employed in his construction: we need to use a fixed point argument to construct the assignment to '$\in$' by approximations when the assignment to '$\rightarrow$' is given. And we need to somehow do these two approximation processes together; this is where most of the difficulties arise.

5.3. Constructing the Valuation of '$\in$': First Steps

OK, let's get down to business. The construction will assign values *in the set* $\{0, \frac{1}{2}, 1\}$ to formulas *relative to two ordinal parameters* α *and* σ (as well as to an assignment of values to the variables in the formula). α will initially be unrestricted; σ will be restricted to being no greater than Ω, the initial ordinal of the cardinality that immediately succeeds that of $|M^+|$. (Forget about Π for now; we will ultimately take it to be at least Ω, but it is not yet in the picture.) We order pairs $\langle \alpha, \sigma \rangle$ lexicographically, that is, $\langle \alpha, \sigma \rangle \preceq \langle \alpha', \sigma' \rangle$ iff either $\alpha < \alpha'$ or both $\alpha = \alpha'$ and $\sigma \leq \sigma'$; the reason for demanding that σ is restricted is so that this defines a genuine sequence. We will mostly be interested in the subsequence of pairs of form $\langle \alpha,\Omega \rangle$; values of σ smaller than Ω serve simply as auxiliaries toward producing the values at Ω. I will call the value of a sentence at the pair $\langle \alpha,\Omega \rangle$ its "value at stage α", and will often drop the Ω from the notation.

I now proceed to assign a value in the set $\{0, \frac{1}{2}, 1\}$ to each formula in L^+ relative to any choice of α, σ and s (the latter being a function assigning objects to the variables); except that as mentioned before I will drop the reference to s by understanding the formulas to be parameterized. I will use the single-bar notation $|A|_{\alpha,\sigma}$ instead of the double-bar notation $\|A\|$ used before, to emphasize that the value space is different. Eventually I will use the two-parameter sequence $|A|_{\alpha,\sigma}$ to recover $\|A\|$. (Just so you know where we're headed, the definition will be that $\|A\|$ is the function whose value at $\alpha < \Pi$ is $|A|_{\Delta+\alpha,\Omega}$; where Δ and Π are ordinals to be specified later. These ordinals will not depend on the particular A; and for all A, $|A|_{\Delta+\Pi,\Omega} = |A|_{\Delta,\Omega}$. Moreover, for all A, $|A|_{\Delta,\Omega}$ is 1 iff for all $\alpha > \Delta$, $|A|_{\alpha,\Omega}$ is 1; and analogously for

0, though not necessarily for $\frac{1}{2}$. These are the main conditions needed to ensure that $\|A\|$ meets the regularity conditions required for membership in the space W^{Π}, which in turn ensures that we get a reasonable logic.)

The single-bar assignment goes as follows:

1. $|o_1 = o_2|_{\alpha,\sigma}$ is 1 if $o_1 = o_2$; 0 otherwise.
2. If p is an atomic predicate of L other than '=', $|p(o_1, \ldots, o_n)|_{\alpha,\sigma}$ is 1 if $\langle o_1, \ldots, o_n\rangle$ is in the extension of p in M; 0 otherwise. (So it's 0 if any of the o_i are in E.)
3. $|\text{Property}(o)|_{\alpha,\sigma}$ is 1 if o is in E; 0 otherwise.
4. $|o_1 \in o_2|_{\alpha,\sigma}$ is 0 if o_2 is in the original domain $|M|$. Otherwise, o_2 is of form $\lambda x \Theta(x, b_1, \ldots, b_n)$ for some specific formula Θ and objects $b_1, \ldots, b_n$. In that case, $|o_1 \in o_2|_{\alpha,\sigma}$ is
   1 if for some $\rho < \sigma$, $|\Theta(o_1, b_1, \ldots, b_n)|_{\alpha,\rho} = 1$;
   0 if for some $\rho < \sigma$, $|\Theta(o_1, b_1, \ldots, b_n)|_{\alpha,\rho} = 0$;
   $\frac{1}{2}$ otherwise.
5. $|\neg A|_{\alpha,\sigma}$ is $1 - |A|_{\alpha,\sigma}$.
6. $|A \wedge B|_{\alpha,\sigma}$ is $\min\{|A|_{\alpha,\sigma}, |B|_{\alpha,\sigma}\}$.
7. $|A \vee B|_{\alpha,\sigma}$ is $\max\{|A|_{\alpha,\sigma}, |B|_{\alpha,\sigma}\}$.
8. $|\forall x A(x)|_{\alpha,\sigma}$ is $\min\{|A(o)|_{\alpha,\sigma} \mid o \in |M^+|\}$.
9. $|\exists x A(x)|_{\alpha,\sigma}$ is $\max\{|A(o)|_{\alpha,\sigma} \mid o \in |M^+|\}$.
10. $|A \rightarrow B|_{\alpha,\sigma}$ is
    1 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $|A|_{\alpha,\Omega} \leq |B|_{\alpha,\Omega}$;
    0 if for some $\beta < \alpha$, and any γ such that $\beta \leq \gamma < \alpha$, $|A|_{\alpha,\Omega} > |B|_{\alpha,\Omega}$;
    $\frac{1}{2}$ otherwise.

Note that when α is held fixed, the values of all atomic predications not involving '$\in$' (including those involving '=' and 'Property'), and of all conditionals, is completely independent of σ: in the case of conditionals, that is because of the use of the specific ordinal Ω on the right hand side of 10. That means that for each fixed value of α we can perform the fixed point construction of [4]. (We perform it "transfinitely many times", once for each α.) More fully, for each α the construction is monotonic in σ: as σ increases with fixed α, the only possible switches in value are from $\frac{1}{2}$ to 0 and from $\frac{1}{2}$ to 1. So by the standard fixed-point argument, the construction must reach a fixed point at some ordinal of cardinality no greater than that of the domain; that is, at some ordinal less than Ω. And that means that we get the following consequence of 4:

(FP) *For all* α *and all* o *and all* Θ *and all* $b_1, \ldots, b_n$, $|o \in \lambda x \Theta(x, b_1, \ldots, b_n)|_{\alpha,\Omega} = |\Theta(o, b_1, \ldots, b_n)|_{\alpha,\Omega}$.

And the rule for the biconditional that follows from 10 and 6 (together with the fact

that an increase in σ stops having any effect by the time we've reached Ω) then implies that for any $\alpha \geq 1$ (and any o, Θ and $b_1, \ldots, b_n$),

$$|o \in \lambda x \Theta(x, b_1, \ldots, b_n) \leftrightarrow \Theta(o, b_1, \ldots, b_n)|_{\alpha,\Omega} = 1;$$

so (dropping Ω from the notation),

(FP-Cor1) *For any* Θ *and* $\alpha \geq 1$, $|\forall u_1 \ldots \forall u_n \exists z \forall x [x \in z \leftrightarrow \Theta(x, u_1, \ldots, u_n)]|_\alpha = 1$.

(FP-Cor1) looks superficially like the (III) that we require, but in fact it falls far short of it, for it says nothing about the double-bar semantic values that we need to guarantee a reasonable logic: nothing about regular functions from Pred(Π) to $\{0, \frac{1}{2}, 1\}$. To do better, we need to explore what happens as we go to higher and higher values of α. That is the goal of the next subsection.

Before proceeding to that, I note a substitutivity result:

(FP-Cor2) *If A is any parameterized formula, and A*results from it by replacing an occurrence of* $y \in \lambda x \Theta(x, o_1, \ldots, o_n)$ *by an occurrence of* $\Theta(y, o_1, \ldots, o_n)$*, then for each* α, $|A|_\alpha = |A^*|_\alpha$.

It's worth emphasizing that this holds even when the substitution is inside the scope of an $\rightarrow$. The proof (whose details I leave to the reader) is an induction on complexity, with a subinduction on α to handle the conditionals and the identity claims. (It is essential that the assignment of values to conditionals for $\alpha = 0$ didn't give conditionals different values when they differ by such a substitution; but it clearly didn't do that, since it gave all conditionals value $\frac{1}{2}$.)

5.4. The Fundamental Theorem

Is there a way to get from our single-bar semantic values relative to levels α to double-bar semantic values in a space W^Π? A naive thought might be to define $\|o_1 \in o_2\|$ as the function that maps each α into $|o_1 \in o_2|_\alpha$. But it should be obvious that this doesn't work: it doesn't meet the regularity condition that we need. (It does work in a few simple cases, like $f_{K \in K}$, but not in general.) The fact that all conditionals have value $\frac{1}{2}$ at $\alpha = 0$ is the most obvious indication of this.

But something like it will work: I will show that there are certain ordinals Δ, which I will call *acceptable ordinals*, with some nice properties. It turns out that if Δ is any acceptable ordinal and Π is any sufficiently larger acceptable ordinal that is also initial (so that it is equal to $\Delta + \Pi$), then we can use this Π for our value space W^Π, and we can define $\|o_1 \in o_2\|$ as the function that maps each $\alpha < \Pi$ into $|o_1 \in o_2|_{\Delta+\alpha}$. The conditions on acceptability will guarantee that the functions are regular. It will also

turn out that even for complex formulas, $\|A\|$ is the function that maps each $\alpha < \Pi$ into $|A|_{\Delta+\alpha}$. And this will guarantee all of the laws that we need.[8]

The definition of acceptability that is easiest to use will require some preliminary explanation. To that end, I introduce a transfinite sequence of functions H_α. (These are the "single-bar analogs of " the Z_α that I informally mentioned in Section 5.2.) H_α is defined as the function that assigns to each parameterized formula A the value $|A|_\alpha$ determined by the single-bar valuation rules. If $v = H_\alpha$, I say that α *represents* v. And if $H_\alpha = H_\beta$ I say that α is *equivalent to* β. I will make use of an intuitively obvious Lemma that the reader can easily prove by induction on γ:

Lemma. *If α is equivalent to β then for any γ, $\alpha + \gamma$ is equivalent to $\beta + \gamma$.*

Now let FINAL be the set of functions v that are represented arbitrarily late, i.e., such that $(\forall\alpha)(\exists\beta \geq \alpha)(v = H_\beta)$.

Proposition 1. FINAL $\neq \emptyset$.

Proof. If it were empty, then for each function v from SENT to $\{0, \frac{1}{2}, 1\}$, there would be an α_v such that $(\forall\beta \geq \alpha_v)(v \neq H_\beta)$. Let θ be the supremum of all the α_v. Then for each function v from SENT to $\{0, \frac{1}{2}, 1\}$, $v \neq H_\theta$. Since H_θ itself is such a function, this is a contradiction. □

Call an ordinal γ *ultimate* if it represents some v in FINAL; that is, if $(\forall\alpha)(\exists\beta \geq \alpha)$ $(H_\gamma = H_\beta)$.

Proposition 2. *If α is ultimate and $\alpha \leq \beta$ then β is ultimate.*

Proof. If $\alpha \leq \beta$, then for some δ, $\beta = \alpha + \delta$. Suppose α is ultimate. Then for any μ, there is an $\eta_\mu \geq \mu$ which is equivalent to α. But then β, i.e., $\alpha + \delta$, is equivalent to $\eta_\mu + \delta$ by the Lemma, and $\eta_\mu + \delta \geq \mu$; so β is ultimate. □

Call a parameterized formula A *ultimately good* if for every ultimate α, $|A|_\alpha = 1$; *ultimately bad* if for every ultimate α, $|A|_\alpha = 0$; and *ultimately indeterminate* if it is neither ultimately good nor ultimately bad. If Γ is a class of parameterized formulas, call an ordinal δ *correct for* Γ if

(ULT) *For any $A \in \Gamma$, $|A|_\delta = 1$ iff A is ultimately good, and $|A|_\delta = 0$ iff A is ultimately bad.*

(It follows that $|A|_\delta = \frac{1}{2}$ iff A is ultimately indeterminate. Also, if Γ is closed under negation then the clause for 0 follows from the clause for 1.) And call an ordinal *acceptable* if it is universally correct, that is, correct for the set of all parameterized formulas. (So if two ordinals are acceptable, they are equivalent, i.e., they assign the same values to every parameterized formula.)

[8] In what follows, I use a slightly different definition of acceptability than in [1], though it is equivalent to the one there; the difference simplifies the proof somewhat.

Proposition 3. *If δ is ultimate, then the following suffices for it to be correct for Γ: for all $A \in \Gamma$, if A is ultimately indeterminate then $|A|_\delta = \frac{1}{2}$.*

Proof. Since δ is ultimate, anything that is ultimately good or ultimately bad has the right value at δ, so only the ultimately indeterminate A have a chance of being treated incorrectly. □

I now proceed to show that there are acceptable ordinals; indeed, arbitrarily large ones. Start with any ultimate ordinal τ, however large. Then every member of FINAL is represented by some ordinal $\geq \tau$; and since FINAL is a set rather than a proper class, and τ is ultimate, there must be a ρ such that $\tau + \rho$ is equivalent to τ and every member of FINAL is represented in the interval $[\tau, \tau + \rho)$. Finally, let Δ be $\tau + \rho \cdot \omega$. I will show that Δ is acceptable.

Proposition 4. *For any n, every member of* FINAL *is represented in the interval* $[\tau + \rho \cdot n, \tau + \rho \cdot (n + 1))$.

Proof. From the fact that $\tau + \rho$ is equivalent to τ, a trivial induction yields that for any finite n, $\tau + \rho \cdot n$ is equivalent to τ; so for any finite n and any $\alpha < \rho$, $\tau + \rho \cdot n + \alpha$ is equivalent to $\tau + \alpha$. So anything represented in the interval $[\tau, \tau + \rho)$ is represented in $[\tau + \rho \cdot n, \tau + \rho \cdot (n + 1))$. □

Proposition 5. Δ *is correct with respect to all conditionals.*

Proof. Since Δ is ultimate, any ultimately good A has value 1 at Δ, and any ultimately bad A has value 0 at Δ. It remains to prove the converses, for the case where A is a conditional.

Suppose $|B \rightarrow C|_\Delta = 1$. Then for some $\alpha < \tau + \rho \cdot \omega$, we have that $(\forall \beta \in [\alpha, \tau + \rho \cdot \omega))(|B|_\beta \leq |C|_\beta)$. Since $\alpha < \tau + \rho \cdot \omega$, there must be an n such that $\alpha < \tau + \rho \cdot n$. So $(\forall \beta \in [\tau + \rho \cdot n, \tau + \rho \cdot \omega))(|B|_\beta \leq |C|_\beta)$. But by Prop. 4, every member of FINAL is represented in $[\tau + \rho \cdot n, \tau + \rho \cdot \omega)$; so for every ultimate ordinal β, $|B|_\beta \leq |C|_\beta$. It follows by the valuation rules that for every ultimate β, $|B \rightarrow C|_\beta = 1$; that is, $B \rightarrow C$ is ultimately good. Similarly, if $|B \rightarrow C|_\Delta = 0$ then $B \rightarrow C$ is ultimately bad. □

Fundamental Theorem: Δ *is acceptable.*

Proof. By Prop. 3, it suffices to show that if A is ultimately indeterminate then $|A|_\Delta = \frac{1}{2}$. Making the mini-stages explicit (and recalling that for any α, if a sentence has value $\frac{1}{2}$ at $\langle \alpha, \Omega \rangle$ then it has that value at all $\langle \alpha, \sigma \rangle$), the claim to be proved is that $(\forall A)(\forall \sigma)$(if $||A|| = \frac{1}{2}$ then $|A|_{\Delta,\sigma} = \frac{1}{2}$). Or reversing the quantifiers, that $(\forall \sigma)(\forall A)$(if $||A|| = \frac{1}{2}$ then $|A|_{\Delta,\sigma} = \frac{1}{2}$). Suppose this fails; let σ_0 be the smallest ordinal at which it fails. We get a contradiction by proving by induction on the complexity of A that

$$(\forall A)(\text{ if } A \textit{ is ultimately indeterminate then } |A|_{\Delta,\sigma_0} = \frac{1}{2}). \qquad (*)$$

If A is atomic with predicate other than '$\in$', then A is not ultimately indeterminate, so the claim is vacuous. Similarly if A is $o_1 \in o_2$ where o_2 is not in E.

Suppose A is $o_1 \in o_2$ where $o_2 \in E$. Then o_2 is $\{x|\Theta(x, b_1, \ldots, b_n)\}$, for some $\Theta(x, b_1, \ldots, b_n)$. So if A is ultimately indeterminate, $\exists x[x = o_1 \wedge \Theta(x, b_1, \ldots, b_n)]$ must be too, since it has the same value as A at each stage. So by choice of σ_0, $|\exists x[x = o_1 \wedge \Theta(x, b_1, \ldots, b_n)]|_{\Delta,\sigma} = \frac{1}{2}$ for all $\sigma < \sigma_0$. But then by the valuation rules, $|o_1 \in o_2|_{\Delta,\sigma_0} = \frac{1}{2}$.

If A is a conditional, then by the valuation rules $|A|_{\Delta,\sigma_0}$ is $|A|_{\Delta,\Omega}$, i.e., $|A|_{\Delta}$, which (when A is ultimately indeterminate) is $\frac{1}{2}$ by Prop. 5.

The other cases use the claim that (*) holds for simpler sentences, and are fairly routine. E.g., if A is $\forall x B$, then if A is ultimately indeterminate, there is a t_0 such that $B(t_0/x)$ is ultimately indeterminate and for no t is $B(t/x)$ ultimately bad. But for any t for which $B(t/x)$ is ultimately indeterminate, including t_0, the induction hypothesis gives that $|B(t_0/x)|_{\Delta,\sigma_0} = \frac{1}{2}$; and for any t for which $B(t/x)$ is ultimately good, $|B(t/x)|_{\Delta,\Omega}$ is 1 and so $|B(t/x)|_{\Delta,\sigma_0} \in \{\frac{1}{2}, 1\}$. So by the valuation rules for $\forall$, $|\forall x B|_{\Delta,\sigma_0} = \frac{1}{2}$. □

5.5. The Valuation of '$\in$' Concluded

We are now ready to choose the value of Π for our space W^{Π}, and to choose a W^{Π}-extension for '$\in$'.

Recall that the acceptable ordinal Δ just constructed was chosen to be bigger than an arbitrarily big τ; so the fundamental theorem gives that acceptable ordinals occur arbitrarily late. Let Δ_0 be the first acceptable ordinal and $\Delta_0 + \delta$ be the second; then an ordinal is acceptable iff it is of form $\Delta_0 + \delta \cdot \beta$.

If I hadn't already imposed stringent requirements on the space of semantic values (so as to be able to develop the semantics generally with as little bother as possible), I could now simply let $\in^*_{o_1 o_2}$ be the function that maps each $\alpha < \delta$ into $|o_1 \in o_2|_{\Delta_0+\alpha}$, and let the set of semantic values be the set of such functions for the different pairs $\langle o_1, o_2\rangle$. But given that I have imposed the stringent requirements, this won't work: I need an acceptable $\Delta_0 + \Pi$ for which Π is an initial ordinal $\geq \Omega$. Also, if I don't insist that Π is strictly greater than δ I will need to prove that for each parameterized formula A there is a ρ_A smaller than δ such that the function $|A|_{\Delta_0+\alpha}$ is ρ_A-cyclic; I imagine that's so, but to avoid taking the trouble to prove it, I will construct Π to be strictly greater than δ, so that we can use δ as a common cycle for all the A.[9]

So let Π be any initial ordinal that is greater than $\Delta_0 + \delta$ and no less than Ω. Since Π is initial, and greater than $\Delta_0 + \delta$, it is identical to $\Delta_0 + \delta \cdot \Pi$, so it is acceptable.

[9] Actually I could avoid a separate stipulation that $\Delta_0 + \delta < \Pi$ by proving this from the stipulation that Π is an initial ordinal, and that is an obvious consequence of what I assume to be a fact, that $\delta < \Delta_0$. But again, there's no need to take the trouble to prove this when an alternative stipulation of the value of Π will obviate the need.

And (since $\Delta_0 + \Pi$ is also just Π), we can carry out the above idea using Π in place of δ:

(E) *For each* σ_1 *and* σ_2, $\in^*_{o_1 o_2}$ *is the function that assigns to each ordinal* $\alpha < \Pi$ *the value* $|o_1 \in o_2|_{\Delta_0+\alpha}$.

Then every value $\|o_1 \in o_2\|$ is δ-cyclic.

The last thing that must be shown, to show that (E) does in fact succeed in assigning a W^Π-extension to '$\in$' for the Π recently chosen, is that each $\|o_1 \in o_2\|$ is regular. But that's clear: if it maps 0 into either 0 or 1, then $|o_1 \in o_2|_{\Delta_0}$ is 0 or 1, so by acceptability, $o_1 \in o_2$ is either ultimately bad or ultimately good, and so $|o_1 \in o_2|_{\Delta_0+\alpha}$ is either 0 for all α or 1 for all α; so $\|o_1 \in o_2\|$ is either $\mathbf{0}$ or $\mathbf{1}$.

So we have a W^Π-model. All that now remains of the consistency proof is to verify that the model validates Axiom Schema (III). This requires the following:

Theorem. *For each parameterized formula* A, $\|A\|$ *is the function that assigns to each ordinal* $\alpha < \Pi$ *the value* $|A|_{\Delta_0+\alpha}$.

Proof. By induction on the complexity of A. It's true by stipulation for membership statements, and trivial for other atomic statements; and the clauses for quantifiers and for connectives other than $\rightarrow$ are completely transparent because the functions assigned these connectives in W^Π-models behave pointwise just like the corresponding connectives behave in the single-bar assignments. This is true for $\rightarrow$ too, except for the behavior at 0. So all we need to verify is the following:

> If $\|A\|$ and $\|B\|$ are the functions that assign to each ordinal $\alpha < \Pi$ the values $|A|_{\Delta_0+\alpha}$ and $|B|_{\Delta_0+\alpha}$ respectively, then $\|A \rightarrow B\|(0)$ assigns the value $|A \rightarrow B|_{\Delta_0}$.

But $\|A \rightarrow B\|(0)$ is by stipulation $(\|A\| \implies \|B\|)(0)$; that is,

> 1 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $\|A\|(\gamma) \leq \|B\|(\gamma)$;
> 0 if for some $\beta < \Pi$, and any γ such that $\beta \leq \gamma < \Pi$, $\|A\|(\gamma) > \|B\|(\gamma)$;
> $\frac{1}{2}$ otherwise.

But $\|A\|(\gamma)$ is by hypothesis $|A|_{\Delta_0+\gamma}$, and likewise for B, so these conditions are just the same as the corresponding conditions for $|A \rightarrow B|_{\Delta_0+\Pi}$. In other words, we've shown that $\|A \rightarrow B\|(0)$ is $|A \rightarrow B|_{\Delta_0+\Pi}$. And since acceptable ordinals are equivalent, that is just $|A \rightarrow B|_{\Delta_0}$, as required. □

Corollary. *Each instance of Axiom Schema* (III) *gets value* **1**.

Proof. We need that for any $o, o_1, \ldots, o_n$,

$$\|o \in \lambda x \Theta(x, o_1, \ldots, o_n)\| = \|\Theta(o, o_1, \ldots, o_n)\|.$$

But by the Theorem, this reduces to the claim that for each α,

$$|o \in \lambda x \Theta(x, o_1, \ldots, o_n)|_{\Delta_0+\alpha} = |\Theta(o, o_1, \ldots, o_n)|_{\Delta_0+\alpha},$$

and that is just a special case of the fixed point result (FP) proved in Section 5.3. □

6. Satisfaction, Sets, and Proper Classes

Without too much trouble, the above construction could be generalized from properties to (non-extensional) n-place relations, for each natural number n. (Properties are the $n = 1$ case. We can include propositions as the $n = 0$ case.) There is a weak way to do this and a strong way. The weak way is to introduce, for each n, the unary predicate 'Rel^n' ('is an n-ary relation') and the $(n + 1)$-place predicate '$\in^n$' (with $x_1, \ldots, x_n \in^n y$ meaning "y is an n-place relation and $\langle x_1, \ldots, x_n \rangle$ instantiates it"); also a single unary predicate 'REL' which each of the 'Rel^n' entail (we need this for restricting the variables to things that aren't relations). The strong way, which requires that the ground language L and ground theory T be adequate to arithmetic and the theory of finite sequences, is to introduce a single binary predicate '$Rel(n, z)$' meaning that z is an n-place relation ('REL' can obviously then be *defined*), and a single binary predicate '$\in$', with '$\in (s, z)$' meaning "for some n, z is an n-place relation and s is an n-place sequence that instantiates z". The details of both the weak and the strong generalization are, as far as I can see, routine. We can also easily build into the language an abstraction symbol that, when applied to any formula $\Theta(x_1, \ldots, x_n, u_1, \ldots, u_k)$ of the language and any k-tuple of entities $o_1, \ldots, o_k$, denotes the n-place relation $\lambda x_1, \ldots, x_n \Theta(x_1, \ldots, x_n, o_1, \ldots, o_k)$; and we can introduce a predicate that applies only to such canonical relations. (In the model we used to prove consistency, all the relations were canonical, but this needn't be so in general.)

From such a generalized theory in the strong form, we could also obtain a consistent theory of expressions and of their satisfaction, a theory that validates the naive schema

$$\langle x_1, \ldots, x_n \rangle \ \textit{satisfies} \ \ulcorner \Theta(v_1, \ldots, v_n) \urcorner \leftrightarrow \Theta(x_1, \ldots, x_n).$$

The basic idea is obvious: identify the formulas of a language that contains a satisfaction predicate with canonical relations, and identify satisfaction with instantiation. Satisfaction claims thus would get values in the space W^{Π}, and excluded middle could not in general be assumed for them. It would be worth being more explicit about the details were it not for the fact that such a theory of satisfaction was given more directly in [1].

A more difficult question is whether we can generalize the above construction to a naive theory of *extensional* relations; or, to stick to the $n = 1$ case, a naive theory of *sets*. Here there do seem to be some difficulties. The matter is a bit complicated because there are several different ways one might propose to treat identity, and there are questions about whether one wants certain laws involving it to hold in full conditional

form or only in the form of rules. But the main problem seems to be independent of these issues, for it doesn't involve identity: the issue is, how can we secure the rule

$$\mathrm{Set}(x) \wedge \mathrm{Set}(y) \wedge \forall w(w \in x \leftrightarrow w \in y) \models \forall z(x \in z \leftrightarrow y \in z),$$

and preferably also the "reverse negated" rule

$$\neg\forall z(x \in z \leftrightarrow y \in z) \models \neg\forall w(w \in x \leftrightarrow w \in y),$$

without any weakening of the Naive Comprehension Schema (III)? The natural way to try to secure these rules is to modify the treatment of '$\in$' so that what the fixed point construction ensures is not the (FP) of Section 5.3, but rather, (FP) only for the special case $\alpha = 0$, supplemented with

(**FP-Mod**) *For all* $\alpha > 0$ *and all* o *and all* Θ *and all* $b_1 \ldots b_n$, $|o \in \lambda x\Theta(x, b_1, \ldots, b_n)|_\alpha = |\exists x[x \equiv o \wedge \Theta(x, b_1, \ldots, b_n)]|_\alpha$,

where '$x \equiv y$' abbreviates '$[\neg\,\mathrm{Set}(x) \wedge x = y] \vee [\mathrm{Set}(x) \wedge \mathrm{Set}(y) \wedge \forall z(z \in x \leftrightarrow z \in y)]$'. (Notice that $|x \equiv y|_\alpha$ depends only on the single-bar values of membership claims for $\beta < \alpha$, given the valuation rules for the biconditional; so there is no threat of circularity.) If we introduce the double-bar values on the basis of the single-bar ones as before, this would yield

$$\|o \in \lambda x\Theta(x, b_1, \ldots, b_n)\| = \|\exists x[x \equiv o \wedge \Theta(x, b_1, \ldots, b_n)]\|.$$

Since $\|o \equiv o\| = \mathbf{1}$ for any o (given that $\equiv$ was defined via $\leftrightarrow$ rather than the material biconditional, and that the single-bar value at $\alpha = 0$ drops out by the time you get to the double-bar values), this would in turn yield

$$\|o \in \lambda x\Theta(x, b_1, \ldots, b_n)\| \succeq \|\Theta(o, b_1, \ldots, b_n)\|,$$

which would ensure the validity of

$$\Theta(o, u_1, \ldots, u_n) \rightarrow o \in \lambda x\Theta(x, u_1, \ldots, u_n). \tag{A}$$

We also get a limited converse, viz., the rule

$$o \in \lambda x\Theta(x, u_1, \ldots, u_n) \models \Theta(o, u_1, \ldots, u_n). \tag{B_1}$$

But this is a significant lessening of naive comprehension: indeed not only do we not get the validity of the conditional

$$o \in \lambda x\Theta(x, u_1, \ldots, u_n) \rightarrow \Theta(o, u_1, \ldots, u_n),$$

we don't even get the "reverse negation" of (B_1), viz.,

$$\neg\Theta(o, u_1, \ldots, u_n) \models o \notin \lambda x\Theta(x, u_1, \ldots, u_n).^{10} \tag{B_2}$$

This does not seem to me enough to count as Naive Set Theory. I don't rule out that we might be able to do better by a more clever construction, but it doesn't look easy.

[10] For a counterexample, let o_1 be $\{w|w \equiv w\}$, o_2 be $\{w|w \equiv w \wedge K \in K\}$ (where K is the Curry set), and o_3 be $\{w|\neg(w \equiv w)\}$; and let $\Theta(x, o_3)$ be '$x \equiv o_3$'. $\|\neg\Theta(o_1, o_3)\| = \mathbf{1}$; but $\|o_1 \notin \lambda x\Theta(x, o_3)\|$ is

But why do we need a naive theory of sets (or other extensional relations) anyway? We have a very nice non-naive theory of sets, namely the Zermelo–Fraenkel theory; and it can be extended to a naive theory of extensional relations either artificially, by defining extensional relations within it by the usual trick, or by a notationally messy but conceptually obvious generalization of ZF that treats multiplace extensional relations autonomously. (Formulations of ZF in terms of a relation of "having no greater rank than" greatly facilitate this generalization.)

It is true that the absence of proper classes in ZF is sometimes awkward. It is also true that adding proper classes in the usual ways (either predicative classes as in Gödel–Bernays, or impredicative ones as in Morse–Kelley) is conceptually unsettling: in each case (and especially in the more convenient Morse–Kelley case) they "look too much like just another level of sets", and the fact that there is no entity that captures the extension of predicates true of proper classes suggests the introduction of still further entities ("super-classes" that can have proper classes as members), and so on *ad infinitum*. But once we have properties (and non-extensional relations more generally), this difficulty is overcome: properties can serve the function that proper classes have traditionally served. The rules they obey are so different from the rules for iterative sets (for instance, they can apply to themselves) that there is no danger of their appearing as "just another level of sets". And since every predicate of properties itself has a corresponding property, there is no fear that the motivation for the introduction of properties will also motivate the introduction of further entities ("super-properties").[11]

Of course, in standard proper class theories, proper classes are extensional; whereas properties are not. Does this show that the properties won't serve the purposes that proper classes have been used for? No. I doubt that extensionality among proper classes plays much role anyway, but without getting into that, one could always use the surrogate $\equiv$ as a "pseudo-identity" among properties that is bound to be adequate in all traditional applications of proper classes; and an extensionality law stated in terms of $\equiv$ rather than = is trivially true. Of course, $\equiv$ is very bad at imitating identity among properties generally: if it weren't, the problem of getting an extensional analog of naive property theory would be easy. But when we confine our attention to those properties that correspond to the proper classes of Gödel–Bernays or Morse–Kelley—in both cases, properties that hold only of things that aren't themselves properties but rather are sets—then $\equiv$ is a very good surrogate for identity: for instance, *over this restricted domain*, excluded middle and all the usual substitutivity principles hold of $\equiv$. Consequently, we have a guarantee that properties will serve all the traditional purposes of proper classes (even in the impredicative Morse–Kelley theory).

My claim, then, is (i) that if we have a naive theory of properties in the background, we have all the advantages of proper classes without the need of any "set-like entities"

$\mathbf{1} - \Upsilon\{\|o \equiv o_1 \wedge \Theta(o, o_3)\|\}$, which is $\preceq \mathbf{1} - \|o_2 \equiv o_1 \wedge \Theta(o_2, o_3)\|$, i.e., $\preceq \mathbf{1} - \|o_2 \equiv o_1 \wedge o_2 \equiv o_3)\|$. But $o_2 \equiv o_1$ has the value $K \in K \rightarrow \top$ and $o_2 \equiv o_3$ has the value $K \in K \rightarrow \bot$; and both assume value $\frac{1}{2}$ at limit ordinals, so $\|o_2 \equiv o_1 \wedge o_2 \equiv o_3)\|$ is not $\mathbf{0}$, so $\|o_1 \notin \lambda x \Theta(x, o_3)\|$ is not $\mathbf{1}$.

[11] The general philosophical view here is quite similar to that in [5], though the theory of properties on offer here is much stronger because of the presence of a serious conditional.

beyond ordinary sets; (ii) that given this naive theory of properties, ordinary iterative set theory (ZF) is a highly satisfactory theory; and (iii) there is no obvious need for any *additional* theory of "naive sets".

But if there is no need of a naive theory of sets, why is there a need for a naive theory of properties, and for a naive theory of satisfaction? Was this paper a wasted effort?

In fact, the case of properties (on at least one conception of them) and of satisfaction are totally different from the case of sets. For the way we solve the paradoxes of naive set theory in ZF is to deny the existence of the alleged set: for instance, there simply is no set of all sets that don't have themselves as members. The analogous paradox in the case of the theory of satisfaction involves the expression 'is not true of itself', and if we were to try to solve the paradox on strictly analogous lines we would have to deny the existence of the expression! That would be absurd: after all, I just exhibited the expression. We could of course say "Sure, there's an expression 'is not true of itself', but it doesn't have the features one would naively think it has, such as being true of just those things that are true of themselves". This would be admitting that the expression exists, but denying the naive satisfaction theory. That is certainly a possible way to go, but it isn't at all like the solution in the ZF case. There are reasons why I don't think it is a *good* way to go: the cost of violating the naive theory of satisfaction is high; see [3]. But without getting into that here, let me simply say that this approach is quite unlike that of ZF (where we deny the existence of the set, instead of saying that it exists but has different members than you might have thought).

The case of properties is slightly more complicated, because there is I believe more than one notion of property. There is, first, the notion of *natural property*, as discussed for instance in [6]. Here we do not want anything like Naive Comprehension: it is central to the idea of natural properties that it is up to science to tell us which natural properties there are. (It is also doubtful that we want natural properties of natural properties. Even if we do, it seems likely that we should adopt a picture which is "ZF-like" in that each natural property has a rank and applies only to non-properties and to properties of lower rank. But there is no need to decide these issues here.) But in addition to the notion of natural property, there is also a conception of property that is useful in semantics. And it is the *raison d'être* of such "semantically conceived properties" (*sc-properties* for short) that every meaningful open sentence (in a given context) corresponds to one.[12] (Open sentences in the language of sc-properties are themselves meaningful, so they must correspond to sc-properties too.) Again, a ZF-like solution in which the existence of the properties is denied goes against the whole point of the notion.

In a theory of semantically conceived properties, then, it is unsatisfactory to say that for a meaningful formula $\Theta(x)$, there is no such thing as $\lambda x \Theta(x)$. It also seems unsatisfactory to say that though $\lambda x \Theta(x)$ exists, the things that instantiate it are not the o for which $\Theta(o)$. In classical logic, those are the only two options, but what I've

[12] Or rather, every meaningful open sentence with a distinguished free variable corresponds to a sc-property relative to any assignment of entities, including possibly sc-properties, to the other free variables.

shown in this paper is how to develop a third option in which we weaken classical logic. If we do that, then we can retain the naive theory of (sc-)properties, and that has an important payoff that has no analogue in the case of sets. At the very least, the value of a naive set theory is unobvious; but the value of a naive theory of satisfaction is overwhelmingly clear, and it is almost as clear that we ought to want a naive theory of sc-properties if we are going to posit sc-properties at all.

You may still want a naive theory of sets, for whatever reason; but what you need is a naive theory of properties and a naive theory of satisfaction. I suspect that you can't get what you want; but you get what you need.

Acknowledgement. This paper first appeared in *The Philosophical Quarterly*, Jan. 2004, ©Basil Blackwell.

References

[1] Field, Hartry: 2003. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic* 32: 139–177.

[2] Field, Hartry: 2003. Is the liar sentence both true and false? In: J. C. Beall and B. Armour-Garb (eds.), *Deflationism and Paradox*, New York: Oxford University Press.

[3] Field, Hartry: 2003. The semantic paradoxes and the paradoxes of vagueness. In: J. C. Beall (ed.), *Liars and Heaps*, New York: Oxford University Press.

[4] Kripke, Saul: 1975. Outline of a theory of truth. *Journal of Philosophy* 72: 690–716.

[5] Maddy, Penelope: 1983. Proper classes. *Journal of Symbolic Logic* 48: 113–39.

[6] Putnam, Hilary: 1975. On properties. In: H. Putnam (ed.), *Mathematics, Matter and Method: Philosophical Papers, vol. 1*, Cambridge: Cambridge University.

NYU Department of Philosophy
503 Silver Center
100 Washington Square East
New York, NY 10003
USA

E-mail: hf18@nyu.edu

The Significance of the Largest and Smallest Numbers for the Oldest Paradoxes

Ulrich Blau

Abstract. Suppose for a moment that the universe **V** of pure sets is as real as anything. Do we know it by axiomatic description? No, *pace* Russell, by acquaintance! For the platonist, **V** is uniquely determined by a formally inexpressible self-referential imperative. Similarly he is acquainted with numbers: A formally inexpressible well-ordered counting process Ω^* leads beyond the finite and transfinite ordinals to transdefinite ordinals without boundary. And conversely, a well-ordered halving process of length Ω^* leads to transdefinite real numbers without boundary. The formal inexpressibility of Ω^* 'Cantor's absolut unendlicher Zahleninbegriff', has significant consequences:

1. the final solution to all semantic paradoxes in a 3-valued logic LR with transdefinite levels of reflexion and an indicator '*' for the reflected level(s),
2. a comprehensive—and maybe complete—theory of formal truth based on the ontology **V** and the logic LR,
3. the formal unsolvability of the vagueness and motion paradoxes which originate from the inconceivable continuum.[1]

The number system S which I will contemplate and commend is Platonist, arch-conservative and leads by natural counting and dividing beyond all largest and smallest numbers. It is characterized by the following.

- S extends the standard models of the ordinal and real numbers, but respects—in contrast to Conway's system [7]—natural properties of the ordinal numbers: There are no magnitudes in S like $\omega - 1$, $\omega/2$, or $\sqrt{\omega}$.
- S combines—for the first time, as far as I know—the natural arithmetics of the largest, medium and smallest numbers.
- The largest numbers one can think of, the *transdefinite ordinals* of S, represent Ω, $\Omega+1, \ldots$ and characterize the well-ordered *counting process*, which transcends the class Ω of all ordinal numbers without boundary.

[1] An extensive treatment of the subject of the present sketch may be found in an unpublished manuscript [5].

- The smallest numbers one can think of, the *transdefinite real numbers* of S, include $\overset{1}{\Omega}$ ('1 over Ω' meaning: '1 2^{Ω}-th', i.e., '1 Ω-th'), $\overset{1}{\Omega+1}$ ('1 over $\Omega+1$'), ... and characterize the well-ordered *halving process*, which transcends the length Ω without boundary.

- The notions of *transdefinite ordinal* and *transdefinite real number* are informally one-place predicates, indeed almost self-evident. Formally, however, these notions, which are necessarily given *relative* to an arbitrary, increasingly extensible well-ordering, cannot be expressed.

This is not without consequences for the oldest paradoxes:

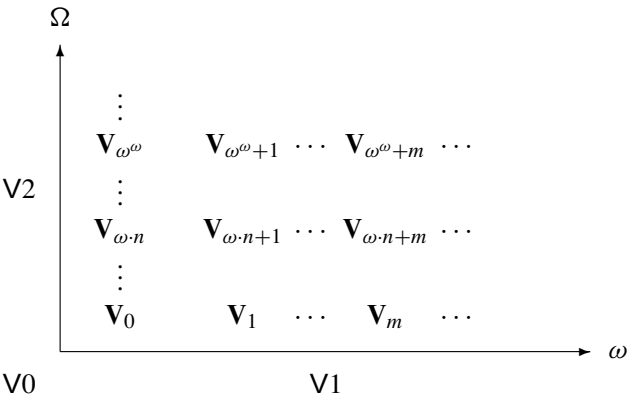
- The notional openness of the transdefinite ordinals allows for a natural non-classical solution to the semantic paradoxes and an understanding of the *whole* formally inexpressible notion of formal truth.

- The notional openness of the transdefinite real numbers prevents any formal understanding of the continuum and the vagueness and motion paradoxes.

We start with three introductory remarks to ontological, semantic and axiomatic prerequisites, which will appeal only to the Platonist. The *ontological basis* of S is the pure universe of sets $\mathbf{V} := \bigcup_{\alpha\in\Omega} \mathbf{V}_\alpha$, where $\mathbf{V}_0 := \emptyset$, $\mathbf{V}_{\alpha+1} := \mathrm{Pot}(\mathbf{V}_\alpha)$, and for limit ordinals λ, $\mathbf{V}_\lambda := \bigcup_{\alpha<\lambda} \mathbf{V}_\alpha$.

$\mathbf{V}$ is extensionally indeterminate as to its height (ordinality) and its width (cardinality). Yet $\mathbf{V}$ is—for the Platonist—in an *intensional and absolute* way *uniquely* and *completely* determined by a formally inexpressible, two-dimensional self-referential imperative:

V0: Consider the empty set $\mathbf{V}_0$!
V1: Consider the power set of the set considered last and repeat V1!
V2: Consider the union of all sets to be considered that way and repeat V1 and V2!



This imperative leads to every rank set $\mathbf{V}_\alpha$. For, if a smallest $\mathbf{V}_\alpha$ stayed out of consideration, then V0, V1 or V2 would be violated, depending on whether α were 0 or a

successor ordinal or a limit ordinal. Therefore, V2 stipulates considering their union $\mathbf{V}$. Then the process of extension goes no further since $\bigcup \mathbf{V} = \mathbf{V}$ and $\mathrm{Pot}(\mathbf{V})$ does not exist, or $\mathrm{Pot}(\mathbf{V}) = \mathbf{V}$ if we somewhat unusually define also for proper classes X:

$$\mathrm{Pot}(X) = \{x \mid x \subseteq X\},$$

with x being a set variable. The *semantic basis* of S is the naturally formulable standard semantics of the impredicative class theory, which is generally not formulated since it raises a seemingly fatal question: What do the class variables range over?—Very simple; over *absolutely all* subclasses of $\mathbf{V}$.—What is the *range* of these X?—None! The subclasses of $\mathbf{V}$ are an incomprehensible plurality; as soon as we look at them as a higher unit, $\mathbf{V}$ being an element (or hyperelement or whatever) thereof, the universal class becomes a set, and the Platonist becomes a relativist. In this paper 'Platonism' means $\mathbf{V}$-absolutism.

As an *axiomatic basis* for S, Bernays' impredicative class theory B suffices, see [1]. To the Platonist, B probably looks *true* in the sense of the standard semantics just mentioned since he is able to justify its essential part, a class-theoretic reflection scheme, from the nature of the universal class:

There is no *specific* statement on $\mathbf{V}$ in any extension of the class-theoretic language by class names: Every truth about $\mathbf{V}$ is reflected in sets $\mathbf{V}_\alpha$ of sufficiently high ranks.

Otherwise, the absolute could be fixed in a formal language, which is unthinkable for the Platonist.

1. The Largest Numbers

Are there numbers that are larger than all large cardinals?[2] Of course there are. Large cardinals are elements of Ω and have *sets* of ordinal predecessors. *Transdefinite ordinals*, on the other hand, are larger than Ω by any amount and have *proper classes* of ordinal predecessors. Let us define the following notions:

A *well-ordering* on a class X is a two-place relation on X which is connected and well-founded.[3] Let W be a well-ordering on X:

A (*proper*) *W-initial segment* of X is a (proper) subclass X_0 of X that contains all W-predecessors of its elements. Then $W_0 := W \cap X_0^2$ is a well-ordering on X_0. W is called a *transdefinite well-ordering* on X if a proper initial segment X_0 of X is

[2] 'Cardinal' in the usual sense. For the Platonist, there are two *transdefinite* cardinalities, encompassing all transdefinite ordinals: the cardinality K of all pure sets and the formally inexpressible cardinality K^* of all proper classes. A formally inexpressible argument shows in an informally cogent way that there are no cardinalities between K and K^*: The class-theoretical continuum hypothesis is *true*! And another informal reflection leads to the truth of the set-theoretical continuum hypothesis, cf. [4].

[3] I do not require, as is widely done, that every proper initial segment of a well-ordering is a set. Otherwise, according to Mostowski's theorem, there would be no transdefinite well-ordering and no transdefinite ordinals.

isomorphic to Ω:

$$(X_0, W_0) \simeq (\Omega, \in).$$

The elements of $X - X_0$ are called *W-transdefinite ordinals*. For every well-ordering W on X there exists a longer well-ordering W^+ on $X^+ = \{\{x\} \mid x \in X\} \cup \{\emptyset\}$: Let

$$W' := \{\langle\{x\}, \{y\}\rangle \mid \langle x, y\rangle \in W\} \text{ and } W^+ := W' \cup \{\langle\{x\}, \emptyset\rangle \mid x \in X\}.$$

W' is isomorphic to W, and W^+ is lengthened by the member $\emptyset$. Thus a longest well-ordering and a largest transdefinite ordinal do not exist any more than the largest natural number. However, while the natural numbers form the set ω and the ordinals form the proper class Ω, the finite, transfinite and transdefinite ordinals are an incomprehensible open plurality. Let us call it—abusively, since any name would be an abuse—Ω^*. Can we *define* Ω^*? Possibly as the 'class of ordinals with respect to *any* well-ordering'? No, with this $\Omega^* = \mathbf{V}$ would hold. But then, which well-ordering would order it? Or as the 'class of members of the *union* of all well-orderings'? No, the heterogeneous well orderings cannot form a union. Can we introduce Ω^* as a basic notion and characterize it axiomatically or semantically? No, we would have to demand:

(A1) Ω^* is well-ordered by a relation W^*.

(A2) Every well-ordering (X, W) is isomorphicto an initial segment of (Ω^*, W^*).

But then it follows, as above, that

$$\Omega^{*+} := \{\{x\} \mid x \in \Omega^*\} \cup \{\emptyset\} \text{ is well ordered by}$$
$$W^{*+} := \{\langle\{x\}, \{y\}\rangle \mid \langle x, y\rangle \in W^*\} \cup \{\langle\{x\}, \emptyset\rangle \mid x \in \Omega^*\},$$

and this well-ordering is not isomorphic to any initial segment of (Ω^*, W^*). Consequently, the most natural extension Ω^* of the most natural mathematical notion—the notion of a natural number—is a formally inexpressible ideal that forces a permanent extension of its own. Any mathematics will have to be content with arbitrary tiny initial segments of the ordinals.

How *could* we know about Ω^*? By the intimate acquaintance of predetermined creation.

P0: Create an arbitrary point!

P1: Add one successor point to the last added point and repeat P1!

P2: Add one successor point to the sequence of all points to be created that way and repeat P1 and P2!

The imperative V0–V2 above must stop at the barrier $\mathbf{V}$. P2 extends the progression of natural numbers created by P0, P1, and will go on forever.

1.1. A Transdefinite Ordinal Segment

This segment of Ω^*, defined as

$$\Omega' := \text{the class of n-tupels of ordinals } \langle \alpha_1, \ldots, \alpha_n \rangle,$$
$$\text{with } \alpha_i \in \Omega, \alpha_1 > 0 \text{ if } n > 1,$$

is a transdefinite generalization of the decimal system. Replacing the small base 10 by the large base Ω results in a correspondence between the 'one-digit' natural numbers $0, \ldots, 9$ and the $\alpha \in \Omega$, and between the 'multiple digit' natural numbers and the transdefinite ordinals of Ω':

$$\langle 1, 0 \rangle, \langle 1, 1 \rangle, \ldots, \langle 1, \alpha \rangle, \ldots, \langle 2, 0 \rangle, \langle 2, 1 \rangle, \ldots, \langle 1, 0, 0 \rangle, \ldots$$

From now on, we shall denote ordinals $\in \Omega'$ by using $\alpha, \beta, \gamma, \ldots$. Ordinals having lengths 1 are called *small* ordinals or just *ordinal numbers*; ordinals having lengths > 1 are called *large* or *transdefinite* ordinals. The well-ordering $<$ of these on Ω' corresponds to the well-ordering of the natural numbers in decimal notation, arranged by the length and, in case of equal lengths of two ordinals, by the first occurrence of a difference: For

$$\alpha = \langle \alpha_1, \ldots, \alpha_m \rangle \text{ and } \beta = \langle \beta_1, \ldots, \beta_n \rangle \text{ we define}$$
$$\alpha < \beta := m < n \vee m = n \wedge \bigvee i (\alpha_i < \beta_i \wedge \bigwedge j < i (\alpha_i = \beta_i)).$$

We classify the ordinals according to the values at their last value places. There are *successor ordinals* with last-place values $\alpha + 1$, *limit ordinals* or *liminals* having limit numbers as last-place values, and 0-*ordinals* which are either 0 itself or large ordinals having last-place values of 0. We call these *litorals*, since they border a cardinally immense ocean, whereas the liminals stand on comparatively firm ground to both sides: immediate successors in front, *close* predecessors behind which are only separated by *sets* of ordinals. (Replacing the last-place value λ of a liminal by any $\beta < \lambda$ results in a close predecessor.) How do we do calculations with these extensionally indeterminate magnitudes?

That is very simple. The well-ordering $<$ on Ω' already determines addition, multiplication and exponentiation in Ω' if the basic principles of arithmetic of natural and ordinal numbers are adhered to:

- *Addition principle*. The *sum* of α and β is the smallest γ which for all $\delta < \beta$ exceeds the sum of α and δ (for $\beta > 0$).
- *Multiplication principle*. The *product* of α and β is the smallest γ which for all $\delta < \beta$ exceeds the product of α and δ by at least α.
- *Exponentiation principle*. The *power* of α and β is the smallest γ which for all $\delta < \beta$ at least exceeds the α-fold power of α and δ (for $\beta > 0$).

These principles, which are independent of notation, are valid for *any* segment of Ω^*. They are, along with the initial equations $\alpha + 0 = \alpha$, $\alpha^0 = 1$, equivalent to the

usual definitions of sum, product and power of natural numbers (restricted to ω) and of ordinal numbers (restricted to Ω). For the segment Ω', we do not consider raising to power for now. We define sum and product as follows. Let L_α be the *length n* of $\alpha = \langle \alpha_1, \ldots, \alpha_n \rangle \in \Omega'$ and $\alpha_i \in \Omega$ be the i^{th} *member* of α.

(D+) If $L_\alpha < L_\beta$, then let $\alpha + \beta := \beta$;
otherwise, if $L_\alpha = m + n,\ L_\beta = n\ (m \geq 0, n \geq 1)$,
then let $\alpha + \beta := \langle \alpha_1, \ldots, \alpha_m, \alpha_{m+1} + \beta_1, \beta_2, \ldots, \beta_n \rangle$

(D·) If $\alpha = 0 \vee \beta = 0$, then let $\alpha \cdot \beta := 0$;
otherwise, if $L_\alpha = m,\ L_\beta = n (m \geq 1, n \geq 1)$,
then let $\alpha \cdot \beta := \langle \beta_1, \ldots, \beta_{n-1}, \alpha_1 \cdot \beta_n, \alpha'_2, \ldots, \alpha'_m \rangle$,
where $\alpha'_i := \begin{cases} \alpha_i & \text{if } \beta_n \text{ is a successor number} \\ 0 & \text{if } \beta_n = 0 \text{ or } \beta_n \text{ is a limit number} \end{cases} \quad (i = 2, \ldots, m)$

This product is curiously entangled, a knot tied by Cantor by switching the natural order of the ordinal factors: '$\alpha \cdot \beta$' means 'α, β times'. If we adapt our writing to the vernacular:

$$\alpha \times \beta := \beta \cdot \alpha \text{ ('}\alpha \text{ times } \beta\text{')}$$

the knot untied:

(D×) If $\alpha = 0$ or $\beta = 0$, then let $\alpha \times \beta := 0$;
otherwise, if $L_\alpha = m,\ L_\beta = n\ (m \geq 1, n \geq 1)$,
then let $\alpha \times \beta := \langle \alpha_1, \ldots, \alpha_{m-1}, \alpha_m \times \beta_1, \beta'_2, \ldots, \beta'_n \rangle$
where $\beta'_i := \begin{cases} \beta_i & \text{if } \alpha_m \text{ is a successor number} \\ 0 & \text{if } \alpha_m = 0 \text{ or } \alpha_m \text{ is a limit number} \end{cases} \quad (i = 2, \ldots, n)$

With these definitions, the addition and multiplication principles are provable; I shall not delve into this here.[4] The main properties of ordinal arithmetic carry over from Ω to Ω', particularly

$$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$$
$$(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$$
$$\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma.$$

Of course, addition and multiplication are not commutative on Ω'; for example, for ordinal numbers $\alpha > 1$ we have:

$$\alpha + \langle 1, 0 \rangle = \langle 1, 0 \rangle < \langle 1, \alpha \rangle = \langle 1, 0 \rangle + \alpha.$$
$$\alpha \cdot \langle 1, 0 \rangle = \langle 1, 0 \rangle < \langle \alpha, 0 \rangle = \langle 1, 0 \rangle \cdot \alpha.$$

[4]The proofs of all statements not proved in the present paper can be found in my unpublished manuscript [5].

How about exponentiation? Generally, $\langle 1,0\rangle^n = \langle 1,0,\ldots,0\rangle$ with n zeros; therefore, $\langle 1,0\rangle^\omega$ is the smallest ordinal outside Ω'. Since $\langle 1,0\rangle$ represents the class Ω in Ω', Ω' is a notational system for the initial segment Ω^ω of Ω^*.

What does a philosopher need transdefinite ordinals for?

1.2. On the Natural Solution to the Semantic Paradoxes

which I can just hint at.[5] It requires a third truth value $\mathcal{O}$ ('open') for semantically unfounded sentences and an omniscient verificater $\mathcal{V}$, who, when confronted with such sentences runs through transdefinite reflexion processes. Let $\mathcal{V}^\alpha$ and $\mathcal{V}^{\geq\alpha}$ denote the omniscient verificater at reflexion level α and at any reflexion level $\geq \alpha$. While trying to evaluate the truth value of the *Liar*,

L: The sentence L is not true

$\mathcal{V}$ is caught in a verification circle, and the value of L remains *open*, as long as $\mathcal{V}^o$ follows the circle without reflecting. $\mathcal{V}^{\geq 1}$ *recognizes* that sentence L is open at level 0. So it is not true, and this is what it claims: $\mathcal{V}^{\geq 1}$ *recognizes* that sentence L is true at level 1. But this is what it denies: $\mathcal{V}^{\geq 2}$ *recognizes* that sentence L is false at level 2. So it is not true, and this is what it claims: $\mathcal{V}^{\geq 3}$ *recognizes* that sentence L is true at level 3 … . And $\mathcal{V}^\omega$ recognizes the infinite oscillation $\mathcal{O}^0, \mathcal{W}^1, \mathcal{F}^2, \mathcal{W}^3, \mathcal{F}^4 \ldots$, of the *Liar* L. In a similar way, he recognizes that

'This sentence is true'	is	$\mathcal{O}^0, \mathcal{F}^1, \mathcal{F}^2, \mathcal{F}^3, \mathcal{F}^4 \ldots,$
'This sentence is false'	is	$\mathcal{O}^0, \mathcal{F}^1, \mathcal{W}^2, \mathcal{F}^3, \mathcal{W}^4 \ldots,$
'This sentence is open'	is	$\mathcal{O}^0, \mathcal{W}^1, \mathcal{F}^2, \mathcal{F}^3, \mathcal{F}^4 \ldots,$
'This sentence is not false'	is	$\mathcal{O}^0, \mathcal{W}^1, \mathcal{W}^2, \mathcal{W}^3, \mathcal{W}^4 \ldots,$
'This sentence is not open'	is	$\mathcal{O}^0, \mathcal{F}^1, \mathcal{W}^2, \mathcal{W}^3, \mathcal{W}^4 \ldots,$

The principle of this is: $\mathcal{V}^\alpha$ recognizes the truth values of all levels $< \alpha$, as well as the values $\mathcal{W}$, $\mathcal{F}$ of founded sentences of his own level α. He neither recognizes the value $\mathcal{O}$ of unfounded sentences of his own level nor values of any higher level. This is the basic idea of the *reflexion logic* LR which is able to solve all paradoxes

[5] It became evident to me some 20 years ago; details are to be found in a meanwhile rather outdated short version [2]. This solution came about independently of existing literature, but is similar to three half-successful attempts: By limiting ourselves to reflexion level 0, we end up at Kripke's proposal [11], the indexical element of truth is implicitly found in Burge [6], and the truth value oscillation appears in Gupta [8].

Compared to these proposals LR (based on set and class theory) has two advantages:

1. the natural semantics: verification trees which model the informal verification and reflexion process;
2. the metatheoretical completeness: every metatheorem for LR is valid in LR on sufficiently high levels of reflection.

concerning truth, satisfaction and denotation in a natural way, formally distinguishing three notions of truth:

$\mathcal{W}$:	unreflected truth, having an invisible indicator for my own level
$\mathcal{W}^*$	reflected truth, having the indicator * for the level(s) I am reflecting now
$\mathcal{W}^\alpha$	truth of the reflexion level α, with constant and variable indices α.

$\mathcal{V}^\alpha$ understands $\mathcal{W}$ as $\mathcal{W}^\alpha$, and $\mathcal{W}^*$ as

- $\mathcal{W}^0$ if $\alpha = 0$;
- $\mathcal{W}^\beta$ if $\alpha = \beta + 1$;
- remaining true from a preceding level up to but excluding α, if α is a liminal or litoral.

LR allows us to formalize sentences like the following, which are then valid from reflexion level ω on:

> The *Liar* L having an unreflected truth predicate $\mathcal{W}$ is *open* at all finite reflexion levels, therefore no paradox arises.

> The *Liar* L having a reflected truth predicate $\mathcal{W}^*$ is *open* at level 0, *true* at all finite odd levels, and *false* at all finite even levels > 0, therefore paradox.

What are transfinite and transdefinite reflexion levels for? They are forced by the following *Liars*:

> This sentence is open at all finite reflexion levels.

It *is*, according to the semantics of LR, open at all finite levels—thus true from level ω on.

> This sentence is open at all transfinite reflexion levels.

It *is*, according to the semantics of LR, open at all transfinite levels—thus true from level Ω, i.e., $\langle 1, 0\rangle$ on. Now:

> This sentence is open at all transdefinite reflexion levels.

Were the ideal concept of a transdefinite ordinal formally expressible *in any way*, the reflexion-logical solution of the semantic paradoxes—the only natural solution as far as one can see—would be as paradoxical as the *Liar ab ovo*.

1.3. The Pure Formal Truth

We call this ideal concept **U**. Its *extension* is the explanandum of the entire logic & mathematics; its *intension*, its rule system, is the central explanandum of logic: the pure formal logic itself. Since **U** cannot be formally expressed, we have to make do with its formal fragments. The three-valued truth predicates $\mathcal{W}^\alpha$ of the hierarchy of reflexion levels are *immeasurably stronger* than the two-valued truth predicates $\mathcal{T}^\alpha$ of the classical hierarchy of object language and metalanguages. Every $\mathcal{T}^\alpha$ (with α in Ω^*) is definable using $\mathcal{W}^0$. Referring with s to the language of set theory we define

$\mathcal{T}_s^\alpha x := \mathcal{W}_s^0 x$, and x is a (Gödel set of a) sentence of the set-theoretical language containing at most truth predicates $\mathcal{T}_s^{\beta<\alpha}$:

This classical hierarchy—let us call it $\mathbf{W}_0$—of set-theoretical truths $\mathcal{T}_s^\alpha$ in **V** is but the *predicative class hierarchy*: The truth predicates $\mathcal{T}_s^\alpha$ with classical levels and the class variables X^α with predicative levels are mutually definable, using the standard interpretation on **V**.
Class-theoretically we define in an analogous manner:

$\mathcal{T}_c^\alpha x := \mathcal{W}_c^0 x$, and x is a (Gödel set of a) sentence of the impredicative-class-theoretical language containing at most truth predicates $\mathcal{T}_c^{\beta<\alpha}$:

Again, the lowest level $\mathcal{T}_c^0$ of impredicative class truths (without truth predicates) is immeasurably stronger than the hierarchy $\mathbf{W}_0$ of predicative class truths or—equivalently—the hierarchy of object language and metalanguages of all set-theoretical truths. This is because the reflexion level hierarchy $\mathbf{W}_1$ of all set-theoretical truths is located between $\mathbf{W}_0$ and $\mathcal{T}_c^0$. Finally, emerging is a four-storey Truth Hierarchy **W**

$\mathbf{W}_3$: three-valued hierarchy of reflected class truths $\mathcal{W}_c^\alpha$

$\mathbf{W}_2$: classical hierarchy of unreflected class truths $\mathcal{T}_c^\alpha$

$\mathbf{W}_1$: three-valued hierarchy of reflected set truths $\mathcal{W}_s^\alpha$

$\mathbf{W}_0$: classical hierarchy of unreflected set truths $\mathcal{T}_s^\alpha$

Each storey has a height of Ω^*. Every truth fragment extends all preceding truth fragments. *If* this three-valued class absolutism is the real, pure theory of truth, then possibly

$$\mathbf{U} = \mathbf{W}$$

holds. In more explicit words: **W** might be extensionally and intensionally correct and complete, i.e., it might *exactly* comprehend all pure formal truths and the logic of the pure formal notion of truth. To surmise this would have hardly been thinkable so far, and it remains unthinkable for the notion of truth in every other science. What is to be said in favour of this conjecture?

For the Platonist, it is evident that **W** is extensionally and intensionally correct. He supposes—after some hesitation—**W** to be complete for the following reasons:

1. To extend **W** ontologically looks impossible to him. If all pure forms are classes $\subseteq$ **V**, there is neither a fifth storey nor a side-building.

2. To make the storeys higher ordinally looks impossible to him since Ω^* cannot be formally reached.

3. Could some hidden aspects or dimensions of truth appear which force us to use further distinctions or indices? This is *a priori* possible, but *a posteriori* improbable, after two and a half millenia of experience with logic. Indeed, the *Liar*, excluded by classical, and then analysed by three-valued levels of truth, was among the earliest logical observations. Other anomalies of the pure formal truth have not come up to the present day.

4. (A delicate point.) Could it be that *means of description* for **U**, **V** or **W** which are not available today will be discovered? This may well be; it may even look highly probable for everyone else. But the Platonist has empirical reasons to be sceptical. He demands that every future basic symbol—like those known—be

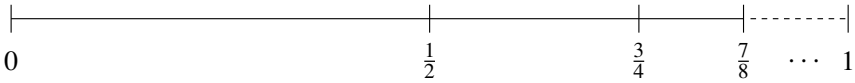
   (a) intensionally unique

   (b) irreducible.

No candidate fulfilling these requirements has emerged in the last hundred years.

2. The Smallest Numbers

Actually, they have been due since the time of Zeno. For everyone has known since then that the interval [0, 1] can be divided into a countably infinite number of partial intervals of decreasing sizes:



and ought to have wondered about a strange gap: [0, 1] is neither by rational nor real numbers x divisible into an infinite number of partial intervals of *equal* sizes, since these numbers are *Archimedean*. For each $x > 0$ there is a natural number n, so that $x \cdot n > 1$. What is missing is a non-Archimedean *infinitesimal* number $\frac{1}{\omega}$ ('1 over ω' meaning: '1 2^{ω}-th', i.e., '1 ω-th') with $\frac{1}{\omega} \cdot \omega = \omega \cdot \frac{1}{\omega} = 1$. Similarly, $\frac{1}{\omega}$ should be

divisible by $2, 4, \ldots\ \omega$, so that

$$\frac{1}{\omega+1}\cdot 2 = 2\cdot\frac{1}{\omega+1} = \frac{1}{\omega},\quad \frac{1}{\omega+2}\cdot 4 = 4\cdot\frac{1}{\omega+2} = \frac{1}{\omega},$$

$$\frac{1}{\omega+\omega}\cdot\omega = \omega\cdot\frac{1}{\omega+\omega} = \frac{1}{\omega},$$

and so on. Thus the desired theory should be an *ordinal* theory. For cardinally, $\omega+1 = \omega+\omega = \omega$ and $\frac{1}{\omega+1} = \frac{1}{\omega+\omega} = \frac{1}{\omega}$ would hold. Our main result, however, will be the special case for transfinite cardinals κ

$$\frac{1}{\kappa}\cdot\kappa = \kappa\cdot\frac{1}{\kappa} = 1,\ \text{and finally}\ \frac{1}{\Omega}\cdot\Omega = \Omega\cdot\frac{1}{\Omega} = 1,$$

since there are no higher κ's. This theory does not seem to exist so far. There is non-standard analysis [12] the models of which contain infinitesimal numbers i and their multiplicatively inverse counterparts j, so that $i\cdot j = j\cdot i = 1$. However, j is not a transfinite ordinal number, but is subject to the standard axioms for real numbers, which are finite.

Somewhat closer to the desired theory are the ingenious numbers of Conway.[6] These are a refinement of real numbers which is transfinite in the ordinal sense - and again satisfy the standard axioms for real numbers. Passionate number crunchers will enjoy them, at the price of having to calculate with numbers like $\omega-1$, $\omega/2$, or $\sqrt{\omega}$, which are larger than all natural numbers but smaller than ω. These numbers do not exist for the Platonist. He takes ω serious as the limit of natural counting: ω has no immediate predecessor $\omega-1$. To put it shortly, for the Platonist, the contemporary understanding of numbers seems to suffer from semantic blindness (of standard mathematics) and technical patchworking (of non-standard mathematics), in a similar way as the understanding of mathematical truth does. This deplorable state is both caused by and results in an operational way of thinking I just caricature for shortness. Mathematics, as the art of calculating, is viewed as ontologically neutral (which is not entirely true, as we shall see), consequently as ontology-free (which is not true at all). Furthermore, without a structurally unique mathematical ontology there is no unique mathematical truth (which is correct). The operationalist, a crisis-proof practitioner, feels in harmony with contemporary spirit. He looks at *number* and *mathematical truth* not as fixed points on the firmament of ideas, but rather as blurred notions with technical variations to be used *ad libitum*. The Platonist is less agile. Though his notion of number is less definite than his notion of mathematical truth, the family of numbers looks to him genetically more closely related than is visible in the number phenotypes on an algebraic level. For the numbers originate at those archaic sources

[6] See [7]. The same holds for the transfinite real numbers of [10] which also satisfy the standard axioms for real numbers. Which kind of numbers are to be viewed as the right ones? Certainly those of Conway and Klaua are more complete algebraically. As to counting and halving, those proposed here are more natural.

of mathematics that are older than calculating and narrower than the broad stream of algebra:

(a) *thinking* in units

(b) *counting* of units

(c) *dividing* of units

(d) *combining* of units to higher level units: sets and classes

(e) *ordering* of units to number-like structures which look natural—and it seems—are found in nature not merely by coincidence.

(a)–(c) is basically the genome of the family of numbers; (b), supported by (d), necessarily leads to the ordinals and cardinals; (c) in a similar way necessarily leads to the real numbers. These theories of the *infinite large* and the *finite small* are rightly considered as final standard theories. What is missing for the Platonist is the conservative unification yielding a single theory of infinitely large and small numbers which respects their nature, e.g., the limit character of ω. Ontology, rather than algebra, leads the way to this theory. We shall extend the standard model of the real numbers, the *binary tree*, by extending its branches to any length and thus reach ever more finely grained and denser transfinite and transdefinite real numbers.

However, we shall not come any closer to the *continuum*—and this refutes the ontological neutrality of mathematics. Zeno's paradox of motion and the paradoxes of vagueness of the *Sorites* (accumulation) type, are formally not solvable since continuity is absolutely without structure and formally incomprehensible: haltless, gliding change without identity and numbering. Every *mathematical* continuum is a fake.

2.1. On the Continuum

If continuity is formally impossible, how and where else then? Every example that comes to mind concerns spatial, temporal or spatiotemporal continuity. So space and time are absolutely structureless? Possibly at the end of the road, certainly not for us. The simplest instance of continuity we can think of is timeless and semiformal: the ideal straight line. It has absolutely no grains. To inquire *how many* points there are on it does not make any more sense than to ask how many angels are dancing on the spire. And now the remarkable, inevitable ontological fall comes about: two ideal straight lines *intersect*. Where? At a *distinct* point. So there are distinct points in the continuum. But the *number* of these points will only be found in representing number spaces. It follows that any mathematization of geometry adulterates the geometric intuition of continuity. This adulteration reaches back farther: *Any* continuity we think of projects *structures* into the structureless and dimensionless continuum. Two intersecting straight lines presuppose a *common plane*—what is the justification for this?

Geometric intuition we shall appeal to in the following sections always aims at a *partial* continuum, e.g., an interval with distinct end points 0, 1. But points, lines, planes and all other appearances between the extremes of pure forms and pure formlessness are invariably a half-understood mixture of projected structure and uncomprehended continuity, substance or materialness which will later dissolve into fine structure—the perennial drama of physics which never frees itself from its subject matter, whichever form it has. Is the *absolute continuum* we stumble on in these material-partial forms at every corner and that determines geometric intuition in such an essential manner, finally the medium between transcendent unity and immanent plurality that Platonists, gnostics and mystics have such a hard time fixing?

Time induces or seduces us to similar speculation. Its continuous elapsing seems evident, almost tautological: no time without elapsing, no elapsing without continuity. The evidence, however, weakens at a closer look. We find ourselves forced to distinguish an internal, mental time from an external, physical time. Does a solid physical body, say Achilles or, more precisely, his centre of gravity, move continuously in the external time? Perhaps. As long as we have no final knowledge about *what exactly* moves in Achilles—quarks? strings? mathematical probability amplitudes?—we shall never understand *how* and *if* anything is really moving there. Possibly every physical motion is a mental artefact.

But the stubbornness of these artefacts, the perceived continuity of motion; aren't they a clear hint at the mental continuity of perception and the continuity—at least in short intervals—of consciousness and internal time? Surely—until we give internal time another look. Intense self-observation at times seems to create discontinuities at the surface of a steady flow of consciousness, at times it seems to imbue continuity therein, and sometimes it approaches absolute emptiness. How much of this is experienced, expected or interpreted is not certain. But one thing is certain: *If* there is real motion or change in the internal or external time, or, *if* elapsing time is not a fundamental deception of consciousness, as some physicists and countless mystics believe, then it implies continuity, which eludes formalization.—What about a total quantization of time to solve all Zenonian problems? This looks like a perfect way out, were it not for the Zenonian problem how a time quantum manages to *elapse*, if absolutely no time passes during its stay. Which cosmic baton will conduct all spacetime quanta to jump synchronously from one state of rest into the next? Still, should future physics convince us that external time and—more difficult—internal time is discontinuous, a last physically inaccessible refuge would remain, *geometric intuition*. The 'continuum' $\mathbb{R}$ of the real numbers owes its discovery to it as well as its undeserved name, since geometric constructions in the ideal Euclidean space are representable precisely, in fact over-precisely, as functions in $\mathbb{R}$.

What *exactly* are real numbers? Geometric intuition does not carry us very far. For example, the Archimedean property mentioned above is geometrically not at all evident. But what is perfectly evident and intensionally unique, up to isomorphism,

is the *binary tree* **B**, which is determined by a formally inexpressible self-referential imperative:[7]

B0: Create an arbitrary point (the 'root')!
B1: Add two successor points to every point that was created in the step performed last and repeat B1!

2.2. Real Numbers

Their most natural model thereof are the branches of **B**, sequences F having lengths ω and values of 0 or 1 at every place. These branches form a (abstract) binary system of notation for the positive real numbers if we interpret the first occurrence of 0 as a *floating point*:

$$F: \quad \underbrace{1, \ldots, 1}_{m \text{ occurrences of } 1}, 0, b_1, \ldots, b_r, \ldots \qquad \text{with } b_i = 0 \text{ or } b_i = 1$$

F denotes the real number $m + b_1/2 + b_2/4 + \ldots b_r/2^r + \ldots$
Some real numbers have two designations, e.g.,

$$11010111\cdots = 11011000\ldots \tag{1}$$

since $2 + 1/2 + 1/8 + 1/16 + 1/32 + \cdots = 2 + 1/2 + 1/4$. We call the sequences on both sides of (1) and generally two sequences F, G with

$$\bigvee n(F(n) = 0 \wedge G(n) = 1 \wedge \bigwedge m < n(F(m) = G(m)) \wedge \bigwedge r > n(F(r) = 1 \wedge G(r) = 0)) \tag{1}$$

left and *right* twins. For a unique notation, we dispense with the left twin. In general, we do away with all *progressive* sequences, i.e., sequences which from a certain place on contain only ones. Now we are able to identify the notation with the positive real

[7] Two things are remarkable for the Platonist in this context. The binary tree **B** is as unique as the progression **P** of natural numbers:

P0: Create an arbitrary point (the 'zero')!
P1: Add one successor point to the last added point and repeat P1!

If **P** is objectively determined, **B** is as well, and the number of branches of **B**, the continuum problem, is objectively determined, see note 1. The other remarkable thing is that the three most important mathematical structures, **P**, **B**, **V**, which are the standard models of the natural number theory (-ies), the real number theory (-ies) and the pure set and class theory (-ies), respectively, can only be determined intensionally, by formally inexpressible self-referential imperatives. Any formal characterization thereof is incomplete or incorrect or semantically circular. It is true that **P** and **B** are categorically characterizable in second-order logic, but the categoricity proof *presupposes* **P** and **B** as objectively determined.

numbers.

$$\mathbb{R}_0 := \text{set of all non-progressive sequences having lengths } \omega \text{ and at every place values of 0 or 1.}$$

A more practical variant, which allows for easier transfinite generalisation, drops the floating point and admits any natural number at the first place.

$$\mathbb{R} := \text{set of all non-progressive sequences having lengths } \omega, \text{ with } F(0) < \omega \text{ and } F(r) = 0 \text{ or } F(r) = 1 \text{ for } r > 0.$$

The *initial value* A_F of F is the natural number $F(0)$, the *fractional value* B_F of F is the infinite sum $F(1)/2+F(2)/4+\ldots F(r)/2^r+\ldots < 1$ (since F is not progressive). From now on we use

i, j, k, m, n	for natural numbers in a metatheoretic, i.e., set-theoretic sense
p, q, r	for natural numbers > 0 in a metatheoretic sense
F, G, R	for real numbers in the sense of $\mathbb{R}$, i.e., sequences in $\mathbb{R}$.

The *order* on $\mathbb{R}$ is given by the first difference:

$$F < G := \bigvee n\,(F(n) < G(n) \wedge \bigwedge m < n\,(F(m) = G(m)))$$

The order is irreflexive, transitive, connected and *continuous*, meaning the following: Every bounded set $M \subseteq \mathbb{R}$ has a supremum $\sup(M)$, i.e., a smallest upper bound $\in \mathbb{R}$

$$(\sup(M))(n) := \bigcup\{F(n) \mid F \in M \wedge \bigwedge m < n\,(F(m) = (\sup(M))(m))\}$$

M is called *bounded* if there is an upper bound G of M, i.e., $\bigwedge F \in M(F \leq G)$. The natural number m is represented in $\mathbb{R}$ by the sequence

$$\overline{m}(n) := \begin{cases} m, & \text{if } n = 0 \\ 0, & \text{otherwise (`}m\text{ at the } 0^{\text{th}} \text{ place, 0 else'}) \end{cases}$$

The rational number $1/2^r$ is represented in $\mathbb{R}$ by the sequence

$$\frac{1}{r}(n) := \begin{cases} 1, & \text{if } n = r \\ 0, & \text{otherwise (`1 at the } r^{\text{th}} \text{ place, 0 else'}) \end{cases}$$

Now,
$A_F := \overline{F(0)}$, the *initial value* of F (in the sense of $\mathbb{R}$) and

$$B_F(n) := \begin{cases} 0, & \text{if } n = 0 \\ F(n), & \text{otherwise, the } \textit{fractional value} \text{ of F (in the sense of } \mathbb{R}). \end{cases}$$

F is called *short*, if

(a) $$\bigvee m \bigwedge n \geq m(F(n) = 0)$$

otherwise, F is called long. The length L_F of F is, informally spoken, the first final 0-place of F, and formally: For short F, L_F is the smallest m, such that (a), for long F, $L_F = \omega$. We adopt the following notation: f, g, h are short numbers $\in \mathbb{R}$. f is either an integer $\overline{m}$ with length 0 or 1 if $m = 0$ or $m > 0$, respectively, or a fractional number with a length > 1 and certain 1-places that are > 0. The last of these 1-places is called the *final place* of f. We denote any fractional f with a final place r by $g^\frown{}_r^1$, where g is the real number that is left when the final place is dropped from f. To economize on the use of parentheses, binding by '$\frown$', '$\cdot$', '$+$' is decreasingly strong. We now define *addition* $F + G$ by induction on L_G using the following rules (+0) through (+3).

(+0) $$(F + \overline{m})(n) := \begin{cases} F(0) + m, & \text{if } n = 0 \\ F(n), & \text{otherwise} \end{cases}$$

Adding $\overline{m}$ just increases the initial value of F by m. Let us now consider the case $F + {}_r^1$. Since ${}_r^1$ represents the number $1/2^r$, we see informally that $F + {}_r^1$ is generated from F as follows

(a) If $F(r) = 0$, then replace this 0 by 1.

(b) If $F(r) = 1$, and F has no 0-place $q < r$ (with $q \geq 1$), then increase the initial value of F by 1 and replace the 1 found in all places 1 to r by 0.

(c) If $F(r) = 1$ and F has a highest 0-place $q < r$, then replace this 0 by 1 and replace the 1 found in all places $q + 1$ to r by 0.

Formally, we state:

If $F(r) = 0$, let

(+1a) $$(F + {}_r^1)(n) := \begin{cases} 1, & \text{if } n = r \\ F(n), & \text{otherwise.} \end{cases}$$

If $F(r) = 1$ and F has no 0-place $q < r$, let

(+1b) $$(F + {}_r^1)(n) := \begin{cases} F(0) + 1, & \text{if } n = 0 \\ 0, & \text{if } 0 < n \leq r \\ F(n), & \text{otherwise.} \end{cases}$$

If $F(r) = 1$ ad F has a highest 0-place $q < r$, let

(+1c) $$(F + \frac{1}{r})(n) := \begin{cases} 1, & \text{if } n = q \\ 0, & \text{if } q < n \leq r \\ F(n), & \text{otherwise.} \end{cases}$$

Using (+0) and (+1), we define for short $G = g^\frown\frac{1}{r}$ by induction on L_G:

(+2) $$F + g^\frown\frac{1}{r} := (F + g) + \frac{1}{r}.$$

The case $g = \overline{0}$ of (+2) is just (+1). Finally, for long G we define:

(+3) $$F + G := \sup\{F + g \mid g < G\}.$$

In the same way we define *multiplication* by induction:

(·0a) $$f \cdot \overline{0} := \overline{0}$$

(·0b) $$f \cdot \overline{m+1} := \overline{m} + f.$$

With this, $f \cdot \frac{1}{r}$ is definable by induction on L_f:

(·1a) $$\overline{m} \cdot \frac{1}{r} := \frac{1}{r} \cdot \overline{m} \text{(the right hand side being defined by (·0))}$$

(·1b) $$f^\frown\frac{1}{q} \cdot \frac{1}{r} := f \cdot \frac{1}{r} + \frac{1}{q+r} \text{ (since } 1/2^q \cdot 1/2^r = 1/2^{q+r}\text{)}$$

(·2) $$f \cdot g^\frown\frac{1}{r} := f \cdot g + f \cdot \frac{1}{r}.$$

The case $g = \overline{0}$ of (·2) is just (·1).

(·3) For long G, let $f \cdot G := \sup\{f \cdot g \mid g < G\}$.

(·4) For long F and arbitrary G, let $F \cdot G := \sup\{f \cdot G \mid f < F\}$.

Subtraction '$\dot{-}$' which is limited on $\mathbb{R}$ and does not yield values below $\overline{0}$, as well as *division* '/' are now definable explicitly:

($\dot{-}$) $$F \dot{-} G := \sup\{h \mid G + h < F\}.$$

(/) For $G > \overline{0}$, let $F/G := \sup\{h \mid G \cdot h < F\}$.

(If $G = \overline{0}$ and $F > \overline{0}$, the set $\{h \mid \overline{0} \cdot h < F\}$ would be an unbounded set $\subseteq \mathbb{R}$ and thus would have no supremum.) Now we consider *negative* real numbers. For the von Neumann numbers 0 and p ($p > 1$), which are used in the metatheory, let

$$0^- := 0 \text{ and } p^- := \{p\}.$$

We extend their linear ordering in a natural way:

$$q^- < p^- < 0^- = 0 < p < q.$$

and assign an *additive inverse* number to every positive real number $F \in \mathbb{R}$:

$$F^- := \{\langle n, i^-\rangle \mid \langle n, i\rangle \in F\}.$$

Furthermore, let $F^{--} = F$. Then

$$\begin{aligned}\mathbb{R}^- &:= \{F^- \mid F \in \mathbb{R}\} \text{ is the set of } \textit{negative real numbers}, \text{ and}\\ \mathbb{R}' &:= \mathbb{R} \cup \mathbb{R}^- \text{ is the set of real numbers } (\mathbb{R} \cap \mathbb{R}^- \text{ being } \{\overline{0}\}).\end{aligned}$$

Again we extend their linear ordering

$$G^- < F^- < \overline{0}^- = \overline{0} < F < G.$$

as well as the four fundamental arithmetical operations, including an extension of restricted substraction '$\dot{-}$' to general substraction '$-$':

(D+)
$$F + G^- := G^- + F := \begin{cases} F \dot{-} G, & \text{if } G \leq F \\ (G \dot{-} F)^-, & \text{if } F < G \end{cases}$$
$$F^- + G^- := (F + G)^-.$$

(D$-$)
$$R - S := R + S^- \text{ for } R, S \in \mathbb{R}'.$$

(D$\cdot$)
$$F \cdot G^- := G^- \cdot F := (F \cdot G)^-.$$
$$F^- \cdot G^- := F \cdot G.$$

(D/)
$$F/G^- := F^-/G := (F/G)^-.$$
$$F^-/G^- := F/G.$$

Instead of $\overline{1}/R$ we shall also write R^{-1}. Now it can be proved that $\mathbb{R}'$ satisfies the standard axioms of real arithmetic.[8] Its formal language contains the classical connectives and first-order quantifiers, the individual constants '0', '1', the relational constant '$<$', the two-place functional constants '+', '$\cdot$', and the one-place functional

[8] This is generally known. However, an exact proof, which I could not find written down anywhere, took almost 20 pages. I would appreciate any communication of a more elegant and complete proof.

constants '$^-$', '$^{-1}$', which are written as upper indices. The axioms are:

(R1) $x + y = y + x$

(R2) $(x + y) + z = x + (y + z)$

(R3) $x \cdot y = y \cdot x$

(R4) $(x \cdot y) \cdot z = x \cdot (y \cdot z)$

(R5) $(x + y) \cdot z = x \cdot z + y \cdot z$

(R6) $x + 0 = x$

(R7) $x \cdot 1 = x$

(R8) $x + x^- = 0$

(R9) $x \neq 0 \to x \cdot x^{-1} = 1$

(R10) $x < y \to \neg y < x$

(R11) $x < y \wedge y < z \to x < z$

(R12) $x < y \vee x = y \vee y < x$

(R13) $x < y \to x + z < y + z$

(R14) $x < y \wedge 0 < z \to x \cdot z < y \cdot z$

(R15) $\bigvee x\, Ax \wedge \bigvee y \bigwedge x\, (Ax \to x \leq y) \to \bigvee y \sup_{Ax} y$

In (R15), let Ax be a first order formula which contains x as a free variable and let $\sup_{Ax}\ y$ be the formula $\bigwedge x\, (Ax \to x \leq y) \wedge \bigwedge z\, (\bigwedge x\, (Ax \to x \leq z) \to y \leq z)$, reads as '$y$ is the supremum of the x for which Ax'. If we substitute the stronger second order axiom

$$\bigvee x\, Xx \wedge \bigvee y \bigwedge x\, (Xx \to x \leq y) \to \bigvee y\, y = \sup(X)$$

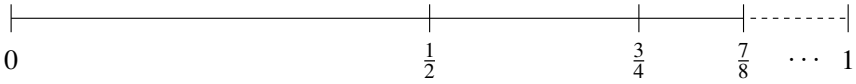
for the first order axiom scheme (R15), the real arithmetic becomes categorical, with all its models being isomorphic to $\mathbb{R}'$. We now refine the positive half $\mathbb{R}$ to create the flexible system S of the largest and smallest numbers.

2.3. Transfinite Real Numbers

Why they have been neglected can only be guessed. To be sure, they are more difficult to handle than real numbers, but not much harder than ordinal numbers. There have probably been two reservations:

- an unjustified distrust of infinite divisibility of the finite:
  (A) $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots = \overline{1}$
- a justified distrust of finite divisibility of the infinite, e.g., of $\overline{\omega} \cdot \frac{1}{1}$ ('half of ω').

On (A). In a metatheoretical sense of $\mathbb{R}$, this equation is correct. The informal sum on the left-hand side is the progressive sequence 0111 . . . , which we view as the left twin-brother of $\overline{1}$. Could it, in extended models $\mathbb{R}_e$, be *infinitesimally smaller* than $\overline{1}$? Quite so. However, (A) is not legitimated and enforced by formal models, but by our *geometric intuition*, the platonic source of mathematics.

(B) 

There is a one-to-one correspondence between the countable infinitely many terms to the left of (A) and the partial intervals of (B). Consequently, the sum corresponds to the entire interval. But if the rational terms $\frac{1}{r}$ $(1 < r < \omega)$ already exhaust the entire interval, what will remain for infinitesimal terms $\frac{1}{\omega}, \frac{1}{\omega+1}, \ldots$? Nothing remains. The infinitesimal terms $\frac{1}{\alpha}$ $(\alpha \geq \omega)$ do not appear in *addition to* the terms $\frac{1}{r}$ $(r < \omega)$. Each $\frac{1}{r}$ rather *contains* arbitrarily large sets and proper classes of infinitesimal terms, transcending $\frac{1}{r}$ in a similar way as the transfinite and transdefinite numbers transcend the finite numbers. Here is a seemingly beautiful argument in favour of (A): If our starting assumption

$$\text{(C)} \qquad \frac{1}{\omega} \cdot \overline{\omega} = \overline{1}$$

is correct, then ω items of the term $\frac{1}{\omega}$ put together yield a value of $\overline{1}$. How could ω terms $\frac{1}{r}$, each of them infinitely larger than $\frac{1}{\omega}$, yield a value of less than $\overline{1}$? Unfortunately, they can. If we drop from ω terms $\frac{1}{r}$ just one, e.g., $\frac{1}{3}$, ω terms still remain, each of them being infinitely larger than $\frac{1}{\omega}$. Yet the sum amounts only to $\overline{7}/\overline{8}$. Experiences like this make (A) suspect, and (C) even more suspect. The only argument favouring (A) I have is geometric intuition: the correspondence between the terms $\frac{1}{r}$ of (A) and *all* partial intervals of (B). As to the fallacious conclusion: Surely we have $\overline{1} - \frac{1}{3} = \overline{7}/\overline{8}$, but $\overline{1} - \frac{1}{\omega}$, i.e., $\iota x\, x + \frac{1}{\omega} = \overline{1}$, does not exist any more than $\overline{\omega} - \overline{1}$, i.e., $\iota x\, x + \overline{1} = \overline{\omega}$ does. For the transfinite real numbers introduced below, $\frac{n}{\omega}$ ('$n\,\omega^{th}s$') divide the interval $[0,\ 1]$ in the same way as the natural numbers n divide the interval $[0,\ \omega]$. As equation (A) halves the interval $[0,\ 1]$ ω times and puts it together again, according to the intuition of geometric continuity it ought to be possible to halve any arbitrarily small interval $[0,\ \frac{1}{\alpha}]$ ω times and put it together again. For this reason we stipulate a generalization of (A):

$$\text{(D)} \qquad \frac{1}{\alpha+1} + \frac{1}{\alpha+2} + \frac{1}{\alpha+3} + \cdots = \frac{1}{\alpha}$$

Much more serious than this is the other subject of reservation, the *finite divisibility of the infinite*. What could $\overline{\omega} \cdot \frac{1}{1}$, 'half of ω', be? Its natural value, $\iota x\, x \cdot \overline{2} = \overline{\omega}$, does not exist: Any $x < \overline{\omega}$ is too small, since this would yield $x \cdot \overline{2} < \overline{\omega}$, and any $x \geq \overline{\omega}$ is too large, since this would yield $x \cdot \overline{2} > \overline{\omega}$ (in the sense of ordinal multiplication). What

else could it be? $\overline{\omega}$ itself looks arithmetically most natural. For, if $\overline{\omega} \cdot \frac{1}{1} < \overline{\omega}$ held, then there would be an n, such that $\overline{\omega} \cdot \frac{1}{1} < \overline{n} \cdot \frac{1}{1}$, and if, on the other hand, $\overline{\omega} \cdot \frac{1}{1} > \overline{\omega}$, then $\overline{\omega} \cdot \frac{1}{1} > \overline{\omega} \cdot \overline{1}$. In both cases, multiplication would not be even weakly monotonic. This is not a sufficient argument yet. The underlying problem is more general: For infinite α and arbitrary β the expression

$$(*) \qquad \overline{\alpha} \cdot \frac{1}{\beta}$$

in most cases has no natural meaning, whereas $\frac{1}{\alpha} \cdot \frac{1}{\beta}$ ('a $2^{\beta-th}$ of a $2^{\alpha-th}$') naturally means $\frac{1}{\alpha+\beta}$ ('a $2^{\alpha+\beta-th}$'). How do we account for this difference? The fundamental intuitions *counting* and *dividing*, applied to their basic object, the *one*, lead to different results. Counting generates units α with structures which determine how to divide them—if they are divisible at all. Division of one generates structureless, continuous units $\frac{1}{\alpha}$, which are further divisible at will. So what can we do? Note that for all $\alpha,\ \beta$ the expression $\frac{1}{\beta} \cdot \overline{\alpha}$, i.e., $\frac{1}{\beta} + \frac{1}{\beta} + \ldots$ (α times in an ordinal sense) does not create any meaning problem. We can count all units we like and sum up them in an ordinal sense. And since for finite $\alpha,\ \beta$

$$\overline{\alpha} \cdot \frac{1}{\beta} = \frac{1}{\beta} \cdot \overline{\alpha},$$

there was no problem in $\mathbb{R}$. The best, at least the simplest solution of the meaning problem $(*)$ seems to be to take $\overline{\alpha} \cdot \frac{1}{\beta}$ as synonymous in general with $\frac{1}{\beta} \cdot \overline{\alpha}$. Then we have, for example

$$\overline{\omega} \cdot \frac{1}{1} = \frac{1}{1} \cdot \overline{\omega} = \frac{1}{1} \cdot (\overline{2} \cdot \overline{\omega}) = \left(\frac{1}{1} \cdot \overline{2}\right) \cdot \overline{\omega} = \overline{1} \cdot \overline{\omega} = \overline{\omega}$$

This solution lacks certain natural divisions of the infinite. For instance, $(\overline{\omega} + \overline{\omega}) \cdot \frac{1}{1}$ has a natural value of $\overline{\omega}$, whereas according to our solution

$$(\overline{\omega} + \overline{\omega}) \cdot \frac{1}{1} = \frac{1}{1} \cdot (\overline{\omega} + \overline{\omega}) = \frac{1}{1} \cdot \overline{\omega} + \frac{1}{1} \cdot \overline{\omega} = \overline{\omega} + \overline{\omega}.$$

It is possible, albeit complicated, to define multiplications which ensure some natural finite divisions of the infinite, at the price of many artificial dividing results. With this in mind, the solution proposed here is incomplete, but correct, if the expressions $\overline{\alpha} \cdot \frac{1}{\beta}$ are simply understood as *defined abbreviations* for $\frac{1}{\beta} \cdot \overline{\alpha}$. Considering perhaps the most important, since nowadays most-violated maxim of logic and philosophy

No *ad hoc* solutions!

and another maxim which follows almost inevitably

Better incomplete than incorrect!

I feel our solution is reasonable.

Before we define the transreal numbers of S, we need some postulates.

(P0) For positive real numbers $x,\ y,\ z$ and their representatives $\overline{x},\ \overline{y},\ \overline{z}$ in S:
$x + y = z \leftrightarrow \overline{x} + \overline{y} = \overline{z}\,,\ x \cdot y = z \leftrightarrow \overline{x} \cdot \overline{y} = \overline{z}.$
For ordinal numbers $\alpha,\ \beta,\ \gamma$ and the representatives $\overline{\alpha},\ \overline{\beta},\ \overline{\gamma}$ in S:
$\alpha + \beta = \gamma \leftrightarrow \overline{\alpha} + \overline{\beta} = \overline{\gamma}\,,\ \alpha \cdot \beta = \gamma \leftrightarrow \overline{\alpha} \cdot \overline{\beta} = \overline{\gamma}.$

This conservation principle will be used unmentioned. Another indispensable principle is the weak monotony of addition and multiplication for all numbers $F,\ G,\ H$:

(P1) $F < G \to H + F \leq H + G \wedge F + H \leq G + H \wedge H \cdot F \leq H \cdot G \wedge$
$F \cdot H \leq G \cdot H.$

The generalization (D) of (A) was

(P2) $$\frac{1}{\alpha+1} + \frac{1}{\alpha+2} + \frac{1}{\alpha+3} + \cdots = \frac{1}{\alpha}$$

Generalizing $\frac{1}{m+n} = \frac{1}{m} \cdot \frac{1}{n}$ we are lead to

(P3) $$\frac{1}{\alpha+\beta} = \frac{1}{\alpha} \cdot \frac{1}{\beta}.$$ – Hence

(P3.1) For finite $A < \overline{1}$ and infinitesimal $\frac{1}{\alpha}$: $A \cdot \frac{1}{\alpha} = \frac{1}{\alpha}$,

since for some n $\frac{1}{n} < A < \overline{1}$, and as $\frac{1}{n} \cdot \frac{1}{\alpha} \overset{(P3)}{=} \frac{1}{n+\alpha} = \frac{1}{\alpha} = \overline{1} \cdot \frac{1}{\alpha}$, (P3.1) follows by (P1). Generalizing $\overline{\alpha} \cdot \frac{1}{\beta} = \frac{1}{\beta} \cdot \overline{\alpha}$ we are lead to

(P4) $\overline{\alpha} \cdot A = A \cdot \overline{\alpha}$, for *fractional* numbers A, i.e., $\overline{0} \leq A < \overline{1}$.

The following associativity is plausible:

(P5) $$\frac{1}{\alpha} \cdot \left(\overline{\beta} \cdot \overline{\gamma}\right) = \left(\frac{1}{\alpha} \cdot \overline{\beta}\right) \cdot \overline{\gamma}$$

since $\frac{1}{\alpha} \cdot \left(\overline{\beta} \cdot \overline{\gamma}\right)$ has the natural meaning

$$\underbrace{\left(\frac{1}{\alpha} + \cdots + \frac{1}{\alpha} + \ldots\right)}_{\beta} + \cdots + \underbrace{\left(\frac{1}{\alpha} + \cdots + \frac{1}{\alpha} + \ldots\right)}_{\beta} + \ldots$$

(a γ-sequence of β-sequences). Hence

(P5.1) For finite $F > 0$ and infinite α: $F \cdot \overline{\alpha} = \overline{\alpha}$,

since for some $n\ \frac{1}{n} < F < \overline{n}$, and as

$$\frac{1}{n} \cdot \overline{\alpha} = \frac{1}{n} \cdot \left(\overline{2^n} \cdot \overline{\alpha}\right) \overset{(P5)}{=} \left(\frac{1}{n} \cdot \overline{2^n}\right) \cdot \overline{\alpha} = \overline{\alpha} = \overline{n} \cdot \overline{\alpha},$$

(P.5.1) follows from (P1). The next associativity, however, does not generally hold.

(P6*) $$\frac{1}{\alpha} \cdot \left(\frac{1}{\beta} \cdot \overline{\gamma}\right) = \left(\frac{1}{\alpha} \cdot \frac{1}{\beta}\right) \cdot \overline{\gamma}.$$

Counterexample:

$$\frac{1}{1} = \frac{1}{1} \cdot \overline{1} \overset{(C)}{=} \frac{1}{1} \cdot \left(\frac{1}{\omega} \cdot \overline{\omega}\right) \neq \left(\frac{1}{1} \cdot \frac{1}{\omega}\right) \cdot \overline{\omega} \overset{(P3)}{=} \frac{1}{1+\omega} \cdot \overline{\omega} = \frac{1}{\omega} \cdot \overline{\omega} \overset{(C)}{=} \overline{1}.$$

Associativity fails because ω absorbs 1 in $1 + \omega$. Therefore we postulate

(P6) $$\frac{1}{\alpha} \cdot \left(\frac{1}{\beta} \cdot \overline{\gamma}\right) = \left(\frac{1}{\alpha} \cdot \frac{1}{\beta}\right) \cdot \overline{\gamma}, \text{ if } \beta \text{ absorbs nothing from } \alpha,$$

more precisely : if α is the smallest α_0 for which

$$\alpha_0 + \beta = \alpha + \beta, \text{ i.e.: if } \bigwedge \alpha_0 < \alpha\ (\alpha_o + \beta < \alpha + \beta).$$

Absorption will be a central feature of transreal arithmetics, and Cantor's normal form, which is unique for all $\alpha > 0$, will be helpful:

(N) $$\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_{r-1}} + \omega^{\alpha_r},\ (\alpha_1 \geq \cdots \geq \alpha_{r-1} \geq \alpha_r \geq 0).$$

ω^{α_1} is the *initial term* $it(\alpha)$, ω^{α_r} is the *final term* $ft(\alpha)$, and $\omega^{\alpha_1} + \cdots + \omega^{\alpha_{r-1}}$ is the *initial sum* $is(\alpha)$ of α. ($is(\alpha) = 0$, if $r = 1$.) Formally:

$$it(\alpha) := \bigcup \{\omega^\beta \mid \omega^\beta \leq \alpha\}$$
$$ft(\alpha) := \bigcap \{\gamma \mid \gamma > 0 \wedge \bigvee \beta\, \alpha = \beta + \gamma\}$$
$$is(\alpha) := \bigcap \{\beta \mid \beta + ft(\alpha) = \alpha\}.$$

For $\alpha = 0$ we define $it(\alpha) = ft(\alpha) = is(\alpha) = 0$. From (N) we get

$$it(\alpha) = \omega^{\alpha_1},\ ft(\alpha) = \omega^{\alpha_r},\ is(\alpha) = \omega^{\alpha_1} + \cdots + \omega^{\alpha_{r-1}}.$$

and generally $\alpha = is(\alpha) + ft(\alpha)$. All additive terms β of the normal form (N) are *additively indecomposable*, i.e., $\bigwedge \alpha < \beta\ (\alpha + \alpha < \beta)$, by a well-known theorem of ordinal arithmetics for $\beta > 0$:

(T1) $$\beta \text{ is additively indecomposable} \leftrightarrow \bigvee \delta\, \beta = \omega^\delta$$
$$\leftrightarrow \bigwedge \alpha < \beta\ (\alpha + \beta = \beta)$$
i.e.,β absorbs all $\alpha < \beta$. Hence

(T2) $$\beta \text{ absorbs nothing from } \alpha \leftrightarrow ft(\alpha) \geq it(\beta).$$

So our previous postulate is

(P6) $$\begin{matrix}1\\ \alpha\end{matrix} \cdot \left(\begin{matrix}1\\ \beta\end{matrix} \cdot \overline{\gamma}\right) = \left(\begin{matrix}1 & 1\\ \alpha & \beta\end{matrix}\right) \cdot \overline{\gamma}, \text{ if } ft(\alpha) \geq it(\beta).$$

Similarly the next postulate has to be restricted.

(P7*) $$\begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\alpha}} = \overline{1}.$$

Some positive examples first. For finite $\alpha = n$, (P7*) gives the correct result $\begin{matrix}1\\ n\end{matrix} \cdot \overline{2^n} = \overline{1}$. Similarly we get our initial conjecture

(C) $$\begin{matrix}1\\ \omega\end{matrix} \cdot \overline{\omega} = \overline{1}, \text{ as } \omega = 2^{\omega} \text{ (in the ordinal sense of } 2^{\omega}.)$$

Furthermore we can infer correctly from (P7*)

$$\begin{matrix}1\\ \omega + \omega\end{matrix} \cdot \overline{2^{\omega+\omega}} = \overline{1},$$

since

$$\begin{matrix}1\\ \omega + \omega\end{matrix} \cdot \overline{2^{\omega+\omega}} \overset{(P3)}{=} \left(\begin{matrix}1 & 1\\ \omega & \omega\end{matrix}\right) \cdot \left(\overline{2^{\omega}} \cdot \overline{2^{\omega}}\right) = \left(\begin{matrix}1 & 1\\ \omega & \omega\end{matrix}\right) \cdot (\overline{\omega} \cdot \overline{\omega})$$
$$\overset{(P5)}{=} \left(\left(\begin{matrix}1 & 1\\ \omega & \omega\end{matrix}\right) \cdot \overline{\omega}\right) \cdot \overline{\omega} \overset{(P6)}{=} \left(\begin{matrix}1\\ \omega\end{matrix} \cdot \left(\begin{matrix}1\\ \omega\end{matrix} \cdot \overline{\omega}\right)\right) \cdot \overline{\omega} \overset{(C)}{=} \begin{matrix}1\\ \omega\end{matrix} \cdot \overline{\omega} \overset{(C)}{=} \overline{1}.$$

Now a counterexample.

$$\begin{matrix}1\\ \omega + 1\end{matrix} \cdot \overline{2^{\omega+1}} = \overline{2},$$

as

$$\begin{matrix}1\\ \omega + 1\end{matrix} \cdot \overline{2^{\omega+1}} \overset{(P3)}{=} \left(\begin{matrix}1 & 1\\ \omega & 1\end{matrix}\right) \cdot \left(\overline{2^{\omega}} \cdot \overline{2}\right) = \left(\begin{matrix}1 & 1\\ \omega & 1\end{matrix}\right) \cdot \left(\overline{\omega} \cdot \overline{2}\right)$$
$$\overset{(P5)}{=} \left(\left(\begin{matrix}1 & 1\\ \omega & 1\end{matrix}\right) \cdot \overline{\omega}\right) \cdot \overline{2} \overset{(P6)}{=} \left(\begin{matrix}1\\ \omega\end{matrix} \cdot \left(\begin{matrix}1\\ 1\end{matrix} \cdot \overline{\omega}\right)\right) \cdot \overline{2} \overset{(P5.1)}{=} \left(\begin{matrix}1\\ \omega\end{matrix} \cdot \overline{\omega}\right) \cdot \overline{2} \overset{(C)}{=} \overline{2}.$$

From these examples one might guess:

(P7**) $$\begin{matrix}1\\ \lambda\end{matrix} \cdot \overline{2^{\lambda}} = \overline{1}, \text{ for all limit ordinals } \lambda.$$

But this leads to contradiction. For instance we would get $\frac{1}{\omega^2+\omega} \cdot \overline{2^{\omega^2+\omega}} = \overline{1}$, and on the other hand we have

$$\begin{aligned}
&\frac{1}{\omega^2+\omega} \cdot \overline{2^{\omega^2+\omega}} \overset{(P3)}{=} \binom{1}{\omega^2} \cdot \binom{1}{\omega} \cdot \overline{2^{\omega+\omega^2+\omega}} = \binom{1}{\omega^2} \cdot \binom{1}{\omega} \cdot \left(\overline{2^{\omega}} \cdot \overline{2^{\omega^2+\omega}}\right) \\
&\overset{(P5)}{=} \left(\binom{1}{\omega^2} \cdot \binom{1}{\omega} \cdot \overline{2^{\omega}}\right) \cdot \overline{2^{\omega^2+\omega}} \overset{(P6)}{=} \left(\binom{1}{\omega^2} \cdot \binom{1}{\omega} \cdot \overline{2^{\omega}}\right)\right) \cdot \overline{2^{\omega^2+\omega}} \\
&\overset{(C)}{=} \frac{1}{\omega^2} \cdot \left(\overline{2^{\omega^2}} \cdot \overline{2^{\omega}}\right) \overset{(P5)}{=} \binom{1}{\omega^2} \cdot \overline{2^{\omega^2}}\Big) \cdot \overline{2^{\omega}} \overset{(P7^{**})}{=} \overline{1} \cdot \overline{2^{\omega}} = \overline{\omega}.
\end{aligned}$$

Let us assume that (P7*) holds for all additively indecomposable α, i.e., by (T1):

(P7) $$\frac{1}{\omega^{\delta}} \cdot \overline{2^{\omega^{\delta}}} = \overline{1}.$$

Now some corollaries will explain our previous examples.

(P7.1) $$\frac{1}{\alpha} \cdot \overline{2^{ft(\alpha)}} = \frac{1}{is(\alpha)},$$

since

$$\begin{aligned}
&\frac{1}{\alpha} \cdot \overline{2^{ft(\alpha)}} = \frac{1}{is(\alpha)+ft(\alpha)} \cdot \overline{2^{ft(\alpha)}} \overset{(P3)}{=} \binom{1}{is(\alpha)} \cdot \binom{1}{ft(\alpha)} \cdot \overline{2^{ft(\alpha)}} \\
&\overset{(P6)}{=} \frac{1}{is(\alpha)} \cdot \binom{1}{ft(\alpha)} \cdot \overline{2^{ft(\alpha)}}\Big) \overset{(P7)}{=} \frac{1}{is(\alpha)} \cdot \overline{1} = \frac{1}{is(\alpha)}.
\end{aligned}$$

(P7.2) $$it(\alpha) = ft(\alpha) \leftrightarrow \frac{1}{\alpha} \cdot \overline{2^{\alpha}} = \overline{1}.$$

'$\rightarrow$': For $\alpha = 0$, since $\frac{1}{0} \cdot \overline{2^0} = \overline{1} \cdot \overline{1} = \overline{1}$; for $\alpha > 0$ let (N) be the normal form of α, and (P7.2) follows by induction on r. The case $r = 1$ is (P7).

In case $r > 1$ we have $\alpha = is(\alpha) + ft(\alpha) = ft(\alpha) + is(\alpha)$, so

$$\begin{aligned}
&\frac{1}{\alpha} \cdot \overline{2^{\alpha}} = \frac{1}{\alpha} \cdot \overline{2^{ft(\alpha)+is(\alpha)}} = \frac{1}{\alpha} \cdot \left(\overline{2^{ft(\alpha)}} \cdot \overline{2^{is(\alpha)}}\right) \overset{(P5)}{=} \binom{1}{\alpha} \cdot \overline{2^{ft(\alpha)}}\Big) \cdot \overline{2^{is(\alpha)}} \\
&\overset{(P7.1)}{=} \frac{1}{is(\alpha)} \cdot \overline{2^{is(\alpha)}} \overset{\text{i.hyp}}{=} \overline{1}.
\end{aligned}$$

– '$\leftarrow$' will follow from

(P7.3) $$it(\alpha) > ft(\alpha) \leftrightarrow \frac{1}{\alpha} \cdot \overline{2^{\alpha}} > \overline{1}.$$

'$\rightarrow$': Then α has a normal form (N) for $r > 1$ and (P7.3) follows by induction r. By (T1) we have $\alpha = ft(\alpha) + \alpha$, therefore

$$\begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\alpha}} = \begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{ft(\alpha)+\alpha}} = \begin{matrix}1\\ \alpha\end{matrix} \cdot \left(\overline{2^{ft(\alpha)}} \cdot \overline{2^{\alpha}}\right) \overset{(P5)}{=} \left(\begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{ft(\alpha)}}\right) \cdot \overline{2^{is(\alpha)+ft(\alpha)}}$$

$$\overset{(P7.1)}{=} \begin{matrix}1\\ is(\alpha)\end{matrix} \cdot \left(\overline{2^{is(\alpha)}} \cdot \overline{2^{ft(\alpha)}}\right) \overset{P5}{=} \left(\begin{matrix}1\\ is(\alpha)\end{matrix} \cdot \overline{2^{is(\alpha)}}\right) \cdot \overline{2^{ft(\alpha)}} \geq \left(\begin{matrix}1\\ is(\alpha)\end{matrix} \cdot \overline{2^{is(\alpha)}}\right) \cdot \overline{2}$$

$$(\text{since } ft(\alpha) \geq 1) \overset{\text{i.hyp \& (P7.2)}}{\geq} \overline{1} \cdot \overline{2} > \overline{1}.$$

As for α either $it(\alpha) = ft(\alpha)$ or $it(\alpha) > ft(\alpha)$, '$\leftarrow$' follows for (P7.2) and (P7.3); therefore

$$\text{(P7.4)} \qquad \begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\alpha}} \geq \overline{1}.$$

$$\text{(P7.5)} \qquad \alpha < \omega^{\delta} \leftrightarrow \begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\omega^{\delta}}} = \overline{2^{\omega^{\delta}}}.$$

'$\rightarrow$': As $\begin{matrix}1\\ \alpha\end{matrix} \leq \overline{1}$, by (P1) $\begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\omega^{\delta}}} \leq \overline{1} \cdot \overline{2^{\omega^{\delta}}} = \overline{2^{\omega^{\delta}}}$. Furthermore we have for $\alpha < \omega^{\delta}$ by (T1) $\alpha + \omega^{\delta} = \omega^{\delta}$, so $\overline{2^{\omega^{\delta}}} = \overline{1} \cdot \overline{2^{\omega^{\delta}}} \overset{(P7.4),(P1)}{\leq} \left(\begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\alpha}}\right) \cdot \overline{2^{\omega^{\delta}}} \overset{(P5)}{=} \begin{matrix}1\\ \alpha\end{matrix} \cdot \left(\overline{2^{\alpha}} \cdot \overline{2^{\omega^{\delta}}}\right) = \begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\alpha+\omega^{\delta}}} \overset{(T1)}{=} \begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\omega^{\delta}}}$. '$\leftarrow$': $\omega^{\delta} \leq \alpha$ implies $\begin{matrix}1\\ \alpha\end{matrix} \leq \begin{matrix}1\\ \omega^{\delta}\end{matrix}$, so $\begin{matrix}1\\ \alpha\end{matrix} \cdot \overline{2^{\omega^{\delta}}} \overset{(P1)}{\leq} \begin{matrix}1\\ \omega^{\delta}\end{matrix} \cdot \overline{2^{\omega^{\delta}}} \overset{(P7)}{=} \overline{1} < \overline{2^{\omega^{\delta}}}$. Some further postulates will be required later.

We now extend the positive real standard model $\mathbb{R} \subseteq {}^{\omega}\omega$ by expanding its sequences to models $\mathbb{R}_{\tau} \subseteq {}^{\tau}\tau$, τ being cardinals $> \omega$. In the following we denote by

$\alpha,\ \beta,\ \gamma,\ \delta$	ordinal numbers $< \tau$
κ	an infinite cardinal number $< \tau$
λ	a limit number $< \tau$
σ	a *leap number*, i.e., 0 or a limit number $< \tau$

The transfinite real numbers of lengths τ, in short, the τ-*real* numbers, will be certain sequences $\in {}^{\tau}\tau$, including $\overline{a}$ and $\begin{matrix}1\\ \alpha\end{matrix}$, which we again define by

$$\overline{\alpha}(\beta) := \begin{cases}\alpha, & \text{if } \beta = 0\\ 0, & \text{otherwise}\end{cases}$$

$$\begin{matrix}1\\ \alpha\end{matrix}(\beta) := \begin{cases}1, & \text{if } \beta = \alpha\\ 0, & \text{otherwise.}\end{cases}$$

$\overline{\alpha}$ represents in $\mathbb{R}_{\tau}$ the ordinal number α, $\begin{matrix}1\\ \alpha\end{matrix}$ represents the value resulting from 1 after the α-halving. Every τ-real number has as *initial value* $F(0)$ an ordinal number $< \tau$. Its *fractional value* is the sum of all $F(\alpha)/2^{\alpha}$ $(0 < \alpha < \tau)$. Successor places $\alpha + 1$ are, as before, occupied by either 1 or 0, depending on whether or not half of $\begin{matrix}1\\ \alpha\end{matrix}$ is a

summation term of the fractional value of F. Values at limit places λ require some consideration. Which sequence F could, for instance, represent the sum $\frac{1}{\omega}+\frac{1}{\omega}$? Since ω has no immediate predecessor, it cannot be $\frac{1}{\omega-1}$. We shall represent $\frac{1}{\omega}+\frac{1}{\omega}$ in $\mathbb{R}_\tau$ by the sequence $\frac{2}{\omega}$ having at the ω^{th} place the value 2 and the value 0 elsewhere. More generally, let $\frac{\alpha}{\beta}$ be the sequence $\in {}^\tau\tau$ for which

$$\frac{\alpha}{\beta}(\gamma) := \begin{cases} \alpha, & \text{if } \gamma = \beta \\ 0, & \text{otherwise.} \end{cases}$$

According to this, all sequences $\overline{\alpha} = \frac{\alpha}{0}$, all sequences $\frac{1}{\alpha+1}$, and for certain $\alpha > 1$, also sequences $\frac{\alpha}{\lambda}$ all belong to $\mathbb{R}_\tau$, the latter meaning

$$\text{(P8)} \qquad \frac{\alpha}{\lambda} = \frac{1}{\lambda} \cdot \overline{\alpha}.$$

A sequence like $\frac{\omega}{\omega}$ will not be admitted, as its value 1 (according to (P8) and (P7)) is already represented by $\overline{1}$. Which values α are admissible at limit places λ? In other words, starting from which height α do double representations arise? The answer follows from (P8) and (P7):

$$\frac{2^{ft(\lambda)}}{\lambda} \overset{\text{(P8)}}{=} \frac{1}{\lambda} \cdot 2^{ft(\lambda)} \overset{\text{(P7.1)}}{=} \frac{1}{is(\lambda)}.$$

$\frac{2^{ft(\lambda)}}{\lambda}$ reduces to $\frac{1}{is(\lambda)}$ and is inadmissible, whereas $\frac{\alpha}{\lambda}$ is irreducible for all $\alpha < 2^{ft(\lambda)}$. So we require

$$F(\lambda) < 2^{ft(\lambda)}$$

in order to avoid double representation. A second requirement for the same purpose results from

$$\text{(P2)} \qquad \frac{1}{\alpha+1} + \frac{1}{\alpha+2} + \frac{1}{\alpha+3} + \cdots = \frac{1}{\alpha}.$$

$\varphi \in {}^\tau\tau$ is called *progressive*, if $\bigvee \alpha \bigwedge n < \omega\, (\varphi(\alpha+n) = 1)$. The infinite sum on the left-hand side is the progressive sequence

$$\varphi(\beta) := \begin{cases} 0, & \text{if } \beta \le \alpha \text{ or } \alpha+\omega \le \beta \\ 1, & \text{if } \alpha < \beta < \alpha+\omega. \end{cases}$$

This is the left twin brother of $\frac{1}{\alpha}$. We drop it. Now we are in a position to define the *set of τ-real numbers*:

$$\mathbb{R}_\tau := \Big\{\varphi \in {}^\tau\tau \mid \bigwedge \alpha < \tau\, (\varphi(\alpha+1) < 2 \wedge (\mathrm{Lim}\,\alpha \to \varphi(\alpha) < 2^{ft(\alpha)})) \\ \wedge \varphi \text{ is not progressive}\Big\}.$$

This can be simplified, since $ft(\alpha+1)=1$ and $2^{ft(\alpha+1)}=2$:

$$\mathbb{R}_\tau := \{\varphi \in {}^\tau\tau \mid \bigwedge \alpha > 0\,(\varphi(\alpha) < 2^{ft(\alpha)}) \wedge \varphi \text{ is not progressive}\}.$$

So $\mathbb{R}_\omega = \mathbb{R}$, and $\mathbb{R}_\tau$ is a most natural generalization. In the following, we denote by $F,\ G,\ H$ τ-real numbers and by $\varphi,\ \psi$ any sequences $\in {}^\tau\tau$. As before, let

$$\varphi < \psi := \bigvee \alpha\,(\varphi(\alpha) < \psi(\alpha) \wedge \bigwedge \beta < \alpha\,(\varphi(\beta) = \psi(\beta))).$$

φ is an *upper bound* of $M \subseteq {}^\tau\tau$, if $\bigwedge \psi \in M\,(\psi \leq \varphi)$. As for the real numbers discussed above we have

Lemma 1. *Every bounded set $M \subseteq {}^\tau\tau$ has a smallest upper bound $\varphi = \sup(M) \in {}^\tau\tau$, i.e., $\varphi(\alpha) := \bigcup\{\psi(\alpha) \mid \psi \in M \wedge \bigwedge \beta < \alpha\,(\psi(\beta) = \varphi(\beta))\}$.*

However, not every bounded set $M \subseteq {}^\tau\tau$ has a smallest upper bound $\in \mathbb{R}_\tau$. To see this, let

$$\mathrm{INF}_\tau := \{F \in \mathbb{R}_\tau \mid \overline{0} < F \wedge \bigwedge n < \omega\,(F(n) = 0)\}$$

be the set of *infinitesimal* τ-real numbers. All $M \subseteq \mathrm{INF}_\tau$ which at some limit place λ cannot be restricted to an $\alpha < 2^{ft(\lambda)}$, i.e., all $M \subseteq \mathrm{INF}_\tau$ for which

$$\bigvee \lambda \bigwedge \alpha < 2^{ft(\lambda)}\,(\bigvee F \in M(F(\lambda) > \alpha))$$

holds have no smallest upper bound $\in \mathbb{R}_\tau$. An example for this is the set INF_τ itself. It is bound by any number $\frac{1}{n}$, but has no smallest upper bound $\in \mathbb{R}_\tau$. $\overline{1}$, however is a 'natural supremum' of INF_τ, meaning the following: For each $F \in \mathrm{INF}_\tau$ there is an n, such that $F < \frac{n}{\omega}$ (e.g., $n = F(\omega)+1$), and for each n there is an $F \in \mathrm{INF}_\tau$, such that $\frac{n}{\omega} < F$ (e.g., $F = \frac{n+1}{\omega}$). By Lemma 1, $\sup(\mathrm{INF}) = \frac{\omega}{\omega}$ is the supremum of INF_τ in ${}^\tau\tau$. Since $\frac{\omega}{\omega}$ has the value $\frac{1}{\omega}\cdot\overline{\omega} = \overline{1}$ we regard 1 as the *natural supremum* of INF_τ *in* $\mathbb{R}_\tau$. Let us call it $\mathrm{Sup}(\mathrm{INF}_\tau)$.

In the same way, we assign a natural supremum $\mathrm{Sup}(M)$ to each bounded set $M \subseteq \mathbb{R}_\tau$, i.e., each set $M \subseteq \mathbb{R}_\tau$ having an upper bound $\in \mathbb{R}_\tau$. We do this in three steps.

Step 1. We take the supremum $\varphi = \sup(M) \in {}^\tau\tau$, which exists according to Lemma 1.

Step 2. We look for the 'smallest natural' $\psi \geq \varphi$ that satisfies the limit place condition for τ-real numbers. If $\varphi(\lambda) < 2^{ft(\lambda)}$ for all λ, then $\psi = \varphi$. Otherwise there is a smallest λ for which $\varphi(\lambda) = 2^{ft(\lambda)}$. By (P8) and (P7.1) $\frac{2^{ft(\lambda)}}{\lambda}$ is reducible to $\frac{1}{is(\lambda)}$. So

$$\psi(\alpha) := \begin{cases} \varphi(\alpha), & \text{if } \alpha < is(\lambda) \\ \varphi(\alpha)+1, & \text{if } \alpha = is(\lambda) \\ 0, & \text{if } \alpha > is(\lambda) \end{cases}$$

is the 'smallest natural' function $> \varphi$ that satisfies the limit place condition.

Step 3. We transform ψ into ψ' which satisfies the condition of non-progressivity. Let for every leap number $\sigma < \tau$

$$\psi_\sigma := \{\langle \sigma + n,\ \psi(\sigma + n)\rangle \mid n < \omega\}$$

be the σ-*section* of ψ. If ψ_σ is not progressive, let $\psi'_\sigma = \psi_\sigma$. Otherwise, there exists a smallest n, such that $\bigwedge r > n\,(\psi(\sigma + r) = 1)$. Then let ψ'_σ be the non-progressive right twin-brother of ψ_σ:

$$\psi'_\sigma(\alpha) := \begin{cases} \psi_\sigma(\alpha), & \text{if } \alpha < \sigma + n \\ 1, & \text{if } \alpha = \sigma + n \\ 0, & \text{if } \alpha > \sigma + n. \end{cases}$$

Now, $\mathrm{Sup}(M) := \bigcup_{\alpha<T} \psi'_\sigma$ is an upper bound ψ^1 for M in $\mathbb{R}_\tau$. We call it the *natural supremum* of M.

α is called a *final place* of F, if $F(\alpha) = \beta > 0$ and for all $\gamma > \alpha\ F(\gamma) = 0$. Again we write.

$$F = G^\frown \begin{matrix}\beta \\ \alpha\end{matrix}$$

where G is the sequence that results if 0 is substituted for β at the final place of F. F is called *closed*, if a final place of F exists, otherwise F is called *open*. The terms 'short' and 'long' previously used for $\mathbb{R}$ are not suitable since G could be open in $F = G^\frown \begin{matrix}\alpha \\ \beta\end{matrix}$. An example for this is

$$F(\alpha) := \begin{cases} 0, & \text{if } \alpha < \omega \text{ and even} \\ 1, & \text{if } \alpha < \omega \text{ and odd} \\ 1, & \text{if } \alpha = \omega \\ 0, & \text{if } \alpha > \omega. \end{cases}$$

Here, ω is the final place of $F = G^\frown \begin{matrix}1 \\ \omega\end{matrix}$, and G is open. The *length* of F is again the first final 0-place of F if existent, otherwise, it is τ.

$$L_F := \min\left(\tau, \bigcap\{\alpha \mid \bigwedge \beta \geq \alpha\,(F(\beta) = 0)\}\right)$$

In the following we denote closed τ-real numbers by $f,\ g,\ h$. It is now possible to define *addition* $F + G$ by induction on L_G, as we did for $\mathbb{R}$, p. 326.

(+0) $$\left(F + \begin{matrix}\beta \\ \sigma\end{matrix}\right)(\alpha) := \begin{cases} F(\sigma) + \beta, & \text{if } \alpha = \sigma \\ F(\alpha), & \text{otherwise.} \end{cases}$$

If $F(\sigma + r) = 0$, then let

(+1a) $$\left(F + \begin{matrix}1 \\ \sigma + r\end{matrix}\right)(\alpha) := \begin{cases} 1, & \text{if } \alpha = \sigma + r \\ F(\alpha), & \text{otherwise.} \end{cases}$$

If $F(\sigma+r)=1$ and F has no 0-place $\sigma+q<\sigma+r$,then let

$$(+1b)\qquad \left(F+\begin{matrix}1\\ \sigma+r\end{matrix}\right)(\alpha):=\begin{cases}F(\sigma)+1, & \text{if } \alpha=\sigma\\ 0, & \text{if } \sigma<\alpha\leq\sigma+r\\ F(\alpha), & \text{otherwise.}\end{cases}$$

If $F(\sigma+r)=1$ and F has a highest 0-place $\sigma+q<\sigma+r$,then let

$$(+1c)\qquad \left(F+\begin{matrix}1\\ \sigma+r\end{matrix}\right)(\alpha):=\begin{cases}1, & \text{if } \alpha=\sigma+q\\ 0, & \text{if } \sigma+q<\alpha\leq\sigma+r\\ F(\alpha), & \text{otherwise.}\end{cases}$$

$$(+2)\qquad F+G^{\frown}\begin{matrix}\beta\\ \alpha\end{matrix}:=(F+G)+\begin{matrix}\beta\\ \alpha\end{matrix}.$$

(+3) For open G let $F+G:=\mathrm{Sup}\{(F+g)\mid g<G\wedge L_g<L_G\}$

Defining multiplication is harder. We write $A,\ B,\ C$ for τ-real fractional numbers, i.e., numbers $<\overline{1}$ Each τ-real number F is of the form $\overline{\alpha}^{\frown}A$, with $\alpha\geq 0$ and $A\geq 0$. We further write $a,\ b,\ c$ for *closed* fractional numbers. Our definition of multiplication is based on some postulates which might appear more evident than they are, for their natural generalizations do not hold:

(1*) $F\cdot G^{\frown}H:=(F\cdot G)+F\cdot H$ – Counterexample:

$$\overline{\omega}\cdot\begin{matrix}1^{\frown}1\\ 1\quad 2\end{matrix}\neq\overline{\omega}\cdot\begin{matrix}1\\ 1\end{matrix}+\overline{\omega}\cdot\begin{matrix}1\\ 2\end{matrix},$$

for

$$\overline{\omega}\cdot\begin{matrix}1^{\frown}1\\ 1\quad 2\end{matrix}\overset{(P4)}{=}\begin{matrix}1^{\frown}1\\ 1\quad 2\end{matrix}\cdot\overline{\omega}\overset{(P5.1)}{=}\overline{\omega}\neq\overline{\omega}+\overline{\omega}\overset{(P5.1)}{=}\begin{matrix}1\\ 1\end{matrix}\cdot\overline{\omega}+\begin{matrix}1\\ 2\end{matrix}\cdot\overline{\omega}\overset{(P4)}{=}\overline{\omega}\cdot\begin{matrix}1\\ 1\end{matrix}+\overline{\omega}\cdot\begin{matrix}1\\ 2\end{matrix}.$$

(2*) $\overline{\alpha}^{\frown}A\cdot F=\overline{\alpha}\cdot F+A\cdot F$ – Counterexample:

$$\overline{1}^{\frown}\begin{matrix}1\\ 1\end{matrix}\cdot\overline{\omega}\neq\overline{1}\cdot\overline{\omega}+\begin{matrix}1\\ 1\end{matrix}+\overline{\omega},$$

for

$$\overline{1}^{\frown}\begin{matrix}1\\ 1\end{matrix}\cdot\overline{\omega}\overset{(P5.1)}{=}\overline{\omega}\neq\overline{\omega}+\overline{\omega}\overset{(P5.1)}{=}\overline{1}\cdot\overline{\omega}+\begin{matrix}1\\ 1\end{matrix}\cdot\overline{\omega}.$$

(3*) $F^{\frown}\begin{matrix}\alpha\\ \beta\end{matrix}=\left(F\cdot\begin{matrix}1\\ \beta\end{matrix}\right)\cdot\overline{\alpha}$ – Counterexample:

$$\overline{\omega} \cdot \underset{\omega}{2} \neq \left(\overline{\omega} \cdot \underset{\omega}{1}\right) \cdot \overline{2},$$

for

$$\overline{\omega} \cdot \underset{\omega}{2} \overset{(P4)}{=} \underset{\omega}{2} \cdot \overline{\omega} \overset{(P8)}{=} \left(\underset{\omega}{1} \cdot \overline{2}\right) \cdot \overline{\omega} \overset{(P5)}{=} \underset{\omega}{1} \left(\overline{2} \cdot \overline{\omega}\right) = \underset{\omega}{1} \cdot \overline{\omega} \overset{(P7)}{=} \overline{1} \neq \overline{2}$$
$$\overset{(P7)}{=} \left(\underset{\omega}{1} \cdot \overline{\omega}\right) \cdot \overline{2} \overset{(P4)}{=} \left(\overline{\omega} \cdot \underset{\omega}{1}\right) \cdot \overline{2}.$$

(4*) $$\underset{\alpha}{\gamma} \cdot \underset{\beta}{1} = \underset{\alpha+\beta}{1} \cdot \overline{\gamma}$$ – Counterexample:

$$\underset{\omega}{2} \cdot \underset{\omega^2}{1} \neq \underset{\omega+\omega^2}{1} \cdot \overline{2},$$

for

$$\underset{\omega}{1} \cdot \underset{\omega^2}{1} \overset{(P3)}{=} \underset{\omega+\omega^2}{1} = \underset{\omega^2}{1} = \overline{1} \cdot \underset{\omega^2}{1}, \text{ and by } \underset{\omega}{1} < \underset{\omega}{2} < \overline{1} \text{ and (P1) follows}$$
$$\underset{\omega}{2} \cdot \underset{\omega^2}{1} = \underset{\omega^2}{1} \neq \underset{\omega^2}{1} \cdot \overline{2} = \underset{\omega+\omega^2}{1} \cdot \overline{2}.$$

The following postulates are weaker than (1*) - (4*)

(P9) $$F \cdot \beta^\frown B = F \cdot \beta + F \cdot B \qquad \text{cf. } (1^*)$$

(P10) $$A \cdot B^\frown \underset{\beta}{\alpha} = A \cdot B + A \cdot \underset{\beta}{\alpha} \qquad \text{cf. } (1^*)$$

(P11) $$\overline{\alpha}^\frown A \cdot B = \overline{\alpha} \cdot B + A \cdot B \qquad \text{cf. } (2^*)$$

(P12) $$A \cdot \underset{\beta}{\alpha} = \left(A \cdot \underset{\beta}{1}\right) \cdot \overline{\alpha} \qquad \text{cf. } (3^*)$$

(P13) $$\underset{\alpha}{\gamma} \cdot \underset{\beta}{1} = \underset{\alpha+\beta}{1} \cdot \overline{\gamma}, \qquad \text{cf. } (4^*)$$

if β absorbs nothing from α, i.e., if $ft(\alpha) \geq it(\beta)$.

Our final postulate needs a similar restriction:

(5*) $$A^\frown \underset{\alpha}{\gamma} \cdot \underset{\beta}{1} = A \cdot \underset{\beta}{1} + \underset{\alpha}{\gamma} \cdot \underset{\beta}{1}.$$

Let us consider 5 examples.

(5a) $$A = \begin{matrix} 1^\frown 1^\frown 1 \\ 1 \quad 2 \quad \omega \end{matrix} \text{ and } \beta = 1.$$

By (5*) follows

(a) $$A \cdot \frac{1}{1} = \frac{1}{1} \cdot \frac{1}{1} + \frac{1}{2} \cdot \frac{1}{1} + \frac{1}{\omega} \cdot \frac{1}{1} \overset{(P3)}{=} \frac{1}{2} + \frac{1}{3} + \frac{1}{\omega + 1}.$$

A natural result: One half of $1/2 + 1/4 + 1/\omega$ is $1/4 + 1/8 + \frac{1}{\omega + 1}$.

(5b) A as above, but $\beta = \omega$. By (P3.1) we get a natural result:

(b) $$A \cdot \frac{1}{\omega} = \frac{1}{\omega}.$$ But by (5*) we get

(b*) $$A \cdot \frac{1}{\omega} = \frac{1}{1} \cdot \frac{1}{\omega} + \frac{1}{2} \cdot \frac{1}{\omega} + \frac{1}{\omega} \cdot \frac{1}{\omega} \overset{(P3)}{=} \frac{1}{\omega} + \frac{1}{\omega} + \frac{1}{\omega + \omega} = \frac{2}{\omega} + \frac{1}{\omega \cdot 2}.$$

An untenable result: $A \cdot \frac{1}{\omega}$ would be greater than $\overline{1} \cdot \frac{1}{\omega}$, although A is smaller than $\overline{1}$, a contradiction to (P1). The reason: In case (a) $\frac{1}{\beta} = \frac{1}{1}$ absorbs none of the three terms $\frac{1}{1}, \frac{1}{2}, \frac{1}{\omega}$ of A, whereas in case (b) $\frac{1}{\beta} = \frac{1}{\omega}$ absorbs the first and greatest term $\frac{1}{1}$, and all the following smaller terms as well.

So we *modify* the distribution of $\frac{1}{\beta}$ over the terms of A: As soon as a first term of A is partly or totally absorbed, all other terms vanish. This gives us the natural result (b) instead of (b*). In the next examples (5c)–(5e) let

$$A = \frac{1}{\omega^\omega} \frown \frac{1}{\omega^\omega + \omega^2} \frown \frac{1}{\omega^\omega + \omega^2 + \omega}.$$

(5c) $\beta = \omega$ or $\omega + r$ or $\omega \cdot r$. In these cases distribution of $\frac{1}{\beta}$ over A absorbs none of the three terms of A partly or totally, and similar to case (a) we have

(c) $$A \cdot \frac{1}{\beta} = \frac{1}{\omega^\omega + \beta} + \frac{1}{\omega^\omega + \omega^2 + \beta} + \frac{1}{\omega^\omega + \omega^2 + \omega + \beta}.$$

(5d) $\beta = \omega^2$. Now $\frac{1}{\beta}$ absorbs none of the two first terms of A, but part of the third, since $\omega + \omega^2 = \omega^2$. So again (c) holds.

(5e) $\beta = \omega^\omega$. Now $\frac{1}{\beta}$ absorbs nothing from the first term of A, but part of the second, since $\omega^2 + \omega^\omega = \omega^\omega$. The third term vanishes, and

(e) $$A \cdot \overset{1}{\beta} = \overset{1}{\omega^\omega + \omega^\omega} + \overset{1}{\omega^\omega + \omega^\omega}.$$

We define:

$\overset{1}{\beta}$ *absorbs* A *at place* $\alpha := \alpha$ is the smallest number for which

$$A(\alpha) > 0 \wedge ft(\alpha) < it(\beta)$$

$\overset{1}{\beta}$ *absorbs nothing from* $A := \bigwedge \alpha\, (A(\alpha) > 0 \rightarrow ft(\alpha) \geq it(\beta))$

and replace (5°∗) by two postulates:

(P14) If $A = A_0 \frown \overset{\gamma}{\alpha} \frown A_1,\ \gamma > 0,$ and $\overset{1}{\beta}$ absorbs A at place α, let

$$A \cdot \overset{1}{\beta} = A_0 \cdot \overset{1}{\beta} + \overset{1}{\alpha + \beta}.$$

(γ vanishes, as for an $\alpha_0 < \alpha :\ \alpha_0 + \beta = \alpha + \beta,$ so $\overset{1}{\alpha_0} \cdot \overset{1}{\beta} = \overset{1}{\alpha} \cdot \overset{1}{\beta}$

and as $\overset{1}{\alpha} \leq \overset{\gamma}{\alpha} < \overset{1}{\alpha_0},$ by (P1) $\overset{\gamma}{\alpha} \cdot \overset{1}{\beta} = \overset{1}{\alpha} \cdot \overset{1}{\beta} = \overset{1}{\alpha + \beta}$.)

(P15) If $A = A_0 \frown \overset{\gamma}{\alpha},\ \gamma > 0,$ and $\overset{1}{\beta}$ absorbs nothing from A, let

$$A \cdot \overset{1}{\beta} = A_0 \cdot \overset{1}{\beta} + \overset{\gamma}{\alpha} \cdot \overset{1}{\beta}.$$

Now *multiplication* $F \cdot G$ is definable by induction on L_F with secondary induction on L_G. Let $a,\ b,\ c$ be *closed fractional numbers*, i.e., numbers < 1 with a final place.

(·0a) $F \cdot \overline{0} := \overline{0}$

(·0b) $F \cdot \overline{\beta + 1} := F \cdot \overline{\beta} + F$

(·0c) $F \cdot \overline{\lambda} := \mathrm{Sup}\{F \cdot \overline{\beta} \mid \beta < \lambda\}$

(·1a) $F \cdot \overline{\beta} \frown B := F \cdot \overline{\beta} + F \cdot B$

(·1b) $\overline{\alpha} \frown A \cdot B := B \cdot \overline{\alpha} + A \cdot B$

(·1c) $A \cdot B \frown \overset{\alpha}{\beta} := A \cdot B + \left(A \cdot \overset{1}{\beta}\right) \cdot \overline{\alpha}$

(·2a) If $A = A_0 \frown \overset{\gamma}{\underset{\alpha}{\frown}} A_1,\ \gamma > 0,$ and $\overset{1}{\underset{\beta}{}}$ absorbs A at place α, let
$$A \cdot \overset{1}{\underset{\beta}{}} := A_0 \cdot \overset{1}{\underset{\beta}{}} + \overset{1}{\underset{\alpha+\beta}{}}.$$

(·2b) If $A = A_0 \frown \overset{\gamma}{\underset{\alpha}{}},\ \gamma > 0,$ and $\overset{1}{\underset{\beta}{}}$ absorbs nothing from A, let
$$A \cdot \overset{1}{\underset{\beta}{}} := A_0 \cdot \overset{1}{\underset{\beta}{}} + \overset{1}{\underset{\alpha+\beta}{}} \cdot \overline{\gamma}.$$

(·2c) If A is open, and $\overset{1}{\underset{\beta}{}}$ absorbs nothing from A, let
$$A \cdot \overset{1}{\underset{\beta}{}} := \mathrm{Sup}\{a \cdot \overset{1}{\underset{\beta}{}} \mid a < A \wedge L_a < L_A\}.$$

(·3) If B is open, let
$$A \cdot B := \mathrm{Sup}\{A \cdot b \mid b < B \wedge L_b < L_B\}.$$

Conditions (·0a)–(·0c), (·2c), (·3) will be obvious and the rest is implied by our postulates: (·1a) is (P9), (·1b) is implied by (P11) and (P4), (·1c) is implied by (P10) and (P12), (·2a) is (P14), (·2b) is implied by (P15) and (P13).

The next step would show that (P0)–(P15) are implied by the definitions of $+$ and $\cdot$, which means that the postulates are consistent relative to ZF. I leave this out and prove just one key postulate.

(P7) $$\overset{1}{\underset{\omega^\delta}{}} \cdot \overline{2^{\omega^\delta}} = \overline{1}.$$

(a) If $\delta = 0$, then $\omega^\delta = 1$, and $\overset{1}{\underset{\omega^\delta}{}} \cdot \overline{2^{\omega^\delta}} = \overset{1}{\underset{1}{}} \cdot \overline{2} = \overline{1}$

(b) If $\delta > 0$, then ω^δ is a limit, and $is(\omega^\delta) = 0$; by (·0) and (+0):

$$\sup\left\{ \overset{1}{\underset{\omega^\delta}{}} \cdot \overline{\beta} \mid \beta < 2^{\omega^\delta} \right\} = \overset{2^{\omega^\delta}}{\underset{\omega^\delta}{}};$$

so for $\varphi = \overset{2^{\omega^\delta}}{\underset{\omega^\delta}{}}$ we have $\varphi(0) = 0$, and $\varphi(0) + 1 = 1$. By definition of the natural supremum follows

$$\mathrm{Sup}\left\{ \overset{1}{\underset{\omega^\delta}{}} \cdot \overline{\beta} \mid \beta < 2^{\omega^\delta} \right\} = \overset{1}{\underset{0}{}} = \overline{1},$$

and by (·0c) follows (P7). This postulate implies our main result for cardinals $\kappa \geq \omega$:

$$\overset{1}{\underset{\kappa}{}} \cdot \kappa = \overline{1},$$

since for $\kappa = \omega : \kappa = \omega^1 = 2^{\omega^1}$, and for $\kappa > \omega : \kappa = \omega^\kappa = 2^{\omega^\kappa}$ (in the sense of ordinal exponentiation.) So the *ordinal halving process* turns out to be a *cardinal dividing process*, and loses the arbitrariness of base 2. $\overset{1}{\underset{\kappa}{}}$ means *one κ-th*, and justifies

the definition

$$\overline{1}/\overline{\kappa} := \frac{1}{\kappa}, \text{ for cardinals } \kappa \geq \omega. \text{ So}$$
$$\overline{1}/\overline{\kappa} \cdot \overline{\kappa} = \overline{\kappa} \cdot \overline{1}/\overline{\kappa} = \overline{1},$$

for the Platonist a mathematical truism, overdue since Zenon (for $\kappa = \omega$) and Cantor (for $\kappa > \omega$).

2.4. Transdefinite Real Numbers

The τ-real numbers are sets: sums of an ordinal $< \tau$ and of fragments generated from the unit *one* by the well-ordered halving process of length $\tau \in \Omega$. Nothing prevents us from continuing this process beyond Ω and from considering transdefinite real numbers which are proper classes.

Now, similar to Ω' of our previous example, let T be any *litoral* initial segment of Ω^*, i.e., for each ordinal it contains, T contains a proper class of larger cardinals. For each $\alpha \in T$ we have $2^\alpha \in T$. Let $\varphi,\ \psi$ be sequences, each having the length T, with values $\in T$. $\varphi < \psi$ is defined as above on p. 325, this time using the well-ordering of T. φ is again called *progressive*, if $\bigvee \alpha \bigwedge n < \omega\, (\varphi(\alpha + n) = 1)$. Similar to p. 333, avoiding the operation $\cap$ which is not applicable in T, we let for $\alpha \in T$

$$ft(\alpha) := \text{ the smallest } \gamma > 0 \text{ such that } \bigvee \beta\, \alpha = \beta + \gamma$$
$$is(\alpha) := \text{ the smallest } \beta \text{ such that } \beta + ft(\alpha) = \alpha$$

We are now ready to define the class predicate (having no extension), in analogy with $\mathbb{R}_\tau$:

φ is a *transdefinite real number of length* T, in short a *T-real number*

$$\mathbb{R}_T\, \varphi := \varphi : T \to T \wedge \bigwedge \alpha > 0\, (\varphi(\alpha) < 2^{ft(\alpha)}) \wedge \varphi \text{ is not progressive.}$$

We define a supremum Sup_P for predicates P of T-real numbers in three steps, in the same way we defined the natural supremum $\mathrm{Sup}(M)$ for sets M of τ-real numbers. P is called *bounded*, if $\bigvee G \bigwedge F\, (P(F) \to F \leq G)$. The first step leads us to $\sup_P$.

Lemma 2. *Every bounded predicate P has a smallest upper bound $\varphi = \sup_P : T \to T$, i.e., $\varphi(\alpha) := \sup\{F(\alpha) \mid P(F) \wedge \bigwedge \beta < \alpha\, (F(\beta) = \varphi(\beta))\}$, being*

$$\sup\{\gamma \mid \bigvee F\, (P(F) \wedge F(\alpha) = \gamma \wedge \bigwedge \beta < \alpha\, (F(\beta) = \varphi(\beta)))\},$$

with $\sup\{\gamma \mid A(\gamma)\}$:= *the smallest δ for which* $\bigwedge \gamma\, (A(\gamma) \to \gamma \leq \delta)$.

Defining φ does not require comprehension of classes to hyperclasses, but rather impredicative class quantification over F. $\varphi = \sup_P$ exists according to a transdefinite induction scheme which is provable in Bernays' class theory B. However, φ is not

always a T-real number. As before, two further steps have to be taken to get the T-real number Sup_P.

Step 2. If $\varphi(\lambda) < 2^{ft(\lambda)}$ for all *limit ordinals*, i.e., liminals and litorals $\lambda \in T$, then let $\psi = \varphi$. Otherwise there exists a smallest λ for which $\varphi(\lambda) = 2^{ft(\lambda)}$, and as above

$$\psi(\alpha) := \begin{cases} \varphi(\alpha), & \text{if } \alpha < is(\lambda) \\ \varphi(\alpha) + 1, & \text{if } \alpha = is(\lambda) \\ 0, & \text{if } \alpha > is(\lambda) \end{cases}$$

is 'the smallest natural' sequence $> \varphi$ satisfying the condition on limit places.

Step 3. As before, this step taking

$$\psi := \bigcup_{\sigma \in T} \psi_\sigma, \text{ generates } \mathrm{Sup}_P := \bigcup_{\sigma \in T} \psi'_\sigma,$$

where $\psi'_\sigma = \psi_\sigma$, or ψ'_σ is the right twin-brother if ψ_σ is progressive. Sup_P is a T-real number, so the definitions of addition and multiplication for $\mathbb{R}_T$ are readily taken from $\mathbb{R}_\tau$. Just two changes have to be kept in mind. λ in (·0c) refers to liminals and litorals. Secondly, Sup is to be understood *predicatively* to avoid hyperclasses. For instance, $\mathrm{Sup}\{F + g \mid g < G \wedge L_g < L_G\}$ means

> the smallest natural upper bound H of all T-real numbers $F + g$ having the property $g < G \wedge L_g < L_G$

(which is for given F, G an impredicative class theoretic description of H.) These definitions of addition and multiplication are valid for all litoral segments T of Ω^*.

Therefore, transreal arithmetic S is invariant against the boundless extensions $\mathbb{R}_T$ of transreal ontology.

So much in short on the largest and smallest numbers. Their guiding ideals, the whole formal truth and the continuum, are formally inexpressible, but not on a par. The former is well-founded in **V**, the latter has neither form nor foundation. Extending and refining our formal net, we shall catch more and more truth fragments, but Achilles and the tortoise will slip through all meshes, and the magic trick, the appearance and disappearance of form and truth in time and consciousness, well remain top secret.

References

[1] Bernays, Paul: 1976. On the problem of schemata of infinity in axiomatic set theory. In: G. H. Müller (ed.), *Sets and Classes. On the Work of Paul Bernays*, Amsterdam: North-Holland, 121–172.

[2] Blau, Ulrich: 1985. Die Logik der Unbestimmtheiten und Paradoxien, Short version. *Erkenntnis* 22: 369–459.

[3] Blau, Ulrich: 1992. Cantor und Nâgârjuna. *Dialectica* 46: 297–311.

[4] Blau, Ulrich: 1998. Ein platonistisches Argument für Cantors Kontinuumshypothese. *Dialectica* 52: 175–202.

[5] Blau, Ulrich: 2000ff. *Die Logik der Unbestimmtheiten und Paradoxien*. Unpublished manuscript, Munich.

[6] Burge, Tyler: 1979. Semantical paradox. *Journal of Philosophy* 76: 169–198.

[7] Conway, John H.: 1976. *On Numbers and Games*. New York: Academic Press.

[8] Gupta, Anil: 1982. Truth and paradox. *Journal of Philosophical Logic* 11: 1–60.

[10] Klaua, Dietrich: 1989. *Transfinite Zahlensysteme auf der Grundlage der Cantorschen Ordinalzahlen*. Karlsruhe: Widmann.

[11] Kripke, Saul A.: 1975. Outline of a theory of truth. *Journal of Philosophy* 72: 690–712.

[12] Robinson, Abraham: 1966. *Non-Standard Analysis*. Amsterdam: North-Holland.

Seminar für Philosophie, Logik und Wissenschaftstheorie
Philosophie-Department
Ludwig-Maximilians-Universität München
Ludwigstr. 31/I
80539 München
Germany

The Prehistory of Russell's Paradox

Nicholas Griffin

Abstract. The paper surveys what is known of the process by which Russell arrived at his paradox, bringing together contributions by Ivor Grattan-Guinness, Gregory H. Moore, Alejandro Garciadiego, and Alberto Coffa. It also addresses the question of why Russell was so disturbed by the result when previous thinkers (Burali-Forti, Cantor), coming upon similar results, had not been.

My purpose in this paper is to trace the line of thought which led to Russell's discovery of his paradox and to his realization that it was, in fact, an important and difficult problem, and in the course of doing this to throw light on a minor puzzle in the historical record. The puzzle is this: Russell was not the first person to find a paradox in set theory, but he was the first to make a really big fuss about it. I want to consider why—both why Russell did and why the others didn't.

So first, the simple facts of chronology. Self-referential paradoxes of the kind that Ramsey classified as semantic or linguistic have been around for a very long time. It's unlikely that St Paul was aware he had created a paradox when he told Titus that a Cretan prophet had truly said that all Cretans are liars, since he used this, to say the least, equivocal evidence as grounds for rebuking them sharply. But writers of greater logical acuity, for example Cervantes, had long used them as jokes and puzzles. There's no particular problem with this. After all, there's no reason to suppose that language is consistent.

Set theory, logic and mathematics are another matter entirely, and it is there that our story properly begins. The chronology here is brief but confused. The story that is told to undergraduates is that the sequence runs

1897:	Burali-Forti's greatest ordinal paradox;
1899:	Cantor's greatest cardinal paradox;
1901:	Russell's class paradox.

This has two advantages: it is easy to remember for exams, and it is brief to expound as an historical digression in a course on set theory. Unfortunately, history is rarely what one would wish it to be, and this is no exception. The actual story is a good deal more complicated and, of the three dates given, only the third is unequivocally correct. Several people have played a part in unravelling the complex historical record, but I'd mention in particular Ivor Grattan-Guinness, Alberto Coffa, Gregory Moore

and Alejandro Garciadiego from whose writings I have learnt a great deal, as will be readily apparent to anyone who knows them.

Consider, first, the greatest ordinal paradox. Let Ω be the ordinal number of the well-ordered set of all ordinals, No. For any ordinal a, $a + 1 > a$. Thus $\Omega + 1 > \Omega$. But, for every $a \in$ No, $a \leq \Omega$. Thus $\Omega + 1 \leq \Omega$. Now, it is clear that an argument along these lines first appeared in [4],[1] but contrary to what van Heijenoort says in reprinting the paper, it did not "immediately arouse the interest of the mathematical world" ([32]: 104). It was rather, as Garciadiego ([19]: 25–26) points out, completely ignored—apart, that is, for a brief note by Burali-Forti himself [5] attempting to correct a misunderstanding in the original paper of Cantor's notion of a well-ordered set.

There is a simpler answer to the question of why the interest of the mathematical world was not immediately aroused by the discovery of Cantor's analogous paradox of the greatest cardinal: Cantor didn't publish it. Nor did he state it—as a paradox. He did, however, note the existence[2] of what he called inconsistent multiplicities of which the set of everything thinkable was one and the set of all cardinals another. He wrote to Dedekind about these in letters of July and August 1899,[3] which establishes a least upper bound on the date. Many earlier dates have been suggested, including 1890, 1892, 1895, and 1896.[4] Cantor himself said[5] he had the idea as early as 1883, but there seems to be no textual evidence for this beyond his recollection. What is clear is that the world could not have learnt of Cantor's inconsistent multiplicities from Cantor himself until 1932 when his letters to Dedekind were published, though the truth is that any mathematician might have discovered them for themselves quite easily once Cantor's power-set theorem was proven [8] by applying the power-set theorem to the set of all sets.

Finally, in May[6] or (early) June 1901[7] comes Russell's paradox. It appeared for the first time in Russell's 1901 draft of Part I of *The Principles of Mathematics* ([*CPBR3*]: 195). It appeared with much greater elaboration and prominence in the final, 1902 draft of the same section of the book (published in 1903). Before the publication of the *Principles* Russell had communicated it to Frege in his famous letter of 16

[1] In fact this is not quite correct: Burali-Forti took No to be not merely well-ordered but perfectly ordered. The matter is discussed in [29].

[2] The word 'existence' is contentious here. See below.

[3] The documentary record even here has been muddied by shoddy editing on Zermelo's part. Two letters, of July and August 1899, were amalgamated into one in his edition of Cantor's *Gesammelte Abhandlungen* [12], see [20]: 126–8. The subsequent edition of the *Briefe* [13] duly separates them again. The partial English translation in ([32]: 113–117), unfortunately, follows Zermelo. There is more detail in Cantor's correspondence with Hilbert, cf. [13]: 390–400. There Cantor calls the sets which are not inconsistent 'ready'.

[4] Evidence for and against all these dates can be assembled—often with difficulty and by inference from brief remarks, see [19]: 33–6.

[5] In a letter of 9 March 1907 to Grace Young, quoted in [29]: 342.

[6] According to Russell's *Autobiography* ([*Auto*]: vol. i, 147).

[7] According to "My Mental Development" (1944), ([*CPBR11*]: 12); "My Debt to German Learning", (1995), ([*CPBR11*]: 107); and a letter to Philip Jourdain, 15 April 1910, see [21]: 132. *My Philosophical Development* ([*MPD*]: 75) says "spring", which suggests not late June.

June 1902 ([32]: 124–5) and to Peano and Whitehead.[8] At about the same time, and quite independently, the group around Hilbert at Göttingen were discovering the paradoxes. When Hilbert received the second volume of Frege's *Grundgesetze* he told Frege that the problem "was known to us here". He credited Zermelo with discovering the Russell paradox "three or four years ago" ([18]: 51). Zermelo had communicated it to Husserl in April 1902 ([28]: 442). According to Grattan-Guinness ([23]: 216) Zermelo mentioned it in print only once ([36]: 191–2). Hilbert included both the Russell-Zermelo and the greatest cardinal paradoxes in his 1905 lecture course on "Logical Principles of Mathematical Thought", see [23]: 216.

Burali-Forti, Cantor, and Russell all had very different attitudes towards these results. As Moore and Garciadiego [29] point out, Burali-Forti was not intending to demonstrate the existence of a contradiction in Cantorian set theory, and did not think he had done so. His explicit aim in the paper was to prove that there were infinite ordinals for which Cantor's trichotomy law fails.[9]

Cantor's case was different again. Burali-Forti at least regarded the contradiction as something to be avoided—by rejecting the proposition from which it was derived, namely the trichotomy law. Cantor, by contrast, saw the contradictions as positive additions to his set theory. In his letters to Dedekind he concludes that both the set of all ordinals and the set of all transfinite cardinals are "absolutely infinite, inconsistent collection[s]" ([11]: 115, 116). There were technical advantages to be gained from accepting these as theorems, as Dauben ([16]: 244) points out.[10] Moreover, Cantor in 1899 realized, as Burali-Forti in 1897 did not, that treating them as a mere step in a *reductio* proof would imperil his whole system.

He also had more philosophical reasons for accepting inconsistent multiplicities. Sets for Cantor were collections whose elements could be regarded as combined into a whole, and which themselves could thus be regarded as a unity. But there were collections, he told Dedekind, such that "the assumption that *all* of [their] elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as 'one finished thing'" ([11]: 114). These were his absolutely infinite, or inconsistent multiplicities, such as the collection of all cardinals, of all thinkable things, and of all classes. Cantor was forced to this distinction by his proofs that every genuine set had a cardinal number and that the power-set of a set always had a greater cardinal number than the original set. If an inconsistent multiplicity, like that of all cardinals, were a genuine set then it should have the greatest possible cardinality. Yet, by the power-set theorem, its power-set should have a greater one. This, for Cantor, was a proof, not that there was something wrong with transfinite set

[8] The letter to Peano is lost. It is not clear whether there ever was a letter to Whitehead—Russell may have communicated the result in person. Whitehead responded by quoting Browning: "never glad confident morning again" ([*MPD*]: 75).

[9] Cantor in fact published a proof of the trichotomy law shortly after Burali-Forti's paper was published ([10]: 216).

[10] The chief of these was that it enabled him to prove the comparability of all cardinals and (so he thought—the proof was in fact mistaken) to show that there were no infinite sets whose powers were not alephs.

theory, but that the collection of all cardinals was not a genuine set, that it had no cardinality, and that there was no greatest cardinal.

It is not, perhaps, surprising that Dedekind found Cantor's inconsistent multiplicities obscure ([20]: 129) or that Robert Bunn denied that Cantor intended to claim that they exist. Bunn writes: "Cantor does not really have 'two kinds' of multiplicities … . Cantor spoke of an inconsistent multiplicity when he might better have spoken of a property such that the supposition that there is a set of all things having that property leads to a contradiction" ([3]: 246). Well, maybe, but this was not at all what Cantor had in mind. For him, inconsistent multiplicities were real, but forever beyond rational comprehension—for to understand them completely would involve, *per impossibile*, treating them as unified, finished things. He linked them ultimately to his religious concerns, see [16]: 245–6.

Cantor's distinction was vague but ultimately achieved mathematical precision in von Neumann's set theory where the distinction between sets and classes is made. There the distinction is drawn between genuine sets, which are elements of other sets, and classes which are not. In von Neumann's set theory the supposition that a class is an element of a set leads to a contradiction. Von Neumann's is a precise, technical distinction but in mathematical intent[11] is not so different from Cantor's. Russell, however, would have been intuitively in tune with Cantor's way of making the distinction, which reflects something of the grandeur of Cantor's original conception of set theory. Russell, as I shall show, would have been sympathetic to the distinction (at least, minus its theological trappings) as he was not towards later consistentizing devices in set theory, to which he always exhibited a profound indifference.

Russell came upon his own paradox while considering Cantor's proof that there is no greatest cardinal. "[I]t seemed to me", he wrote, "that the number of all the things in the world ought to be the greatest possible".[12] "Accordingly", he continues, "I examined his proof with some minuteness, and endeavoured to apply it to the class of all things there are. This led me to consider those classes which are not members of themselves, and to ask whether the class of such classes is or is not a member of itself." ([*Auto*]: vol. i, 147, see also [*POM*]: 362n.). It was Russell who saw the resulting paradox as a serious problem to be solved. But not immediately: at first he thought "there was some trivial error in the reasoning" (*ibid.*). This, no doubt, accounts for his relatively long silence on the topic. Though the paradox was discovered in May/June 1901, Russell does not seem to have reported it to anyone (except perhaps Whitehead) until June 1902, when he wrote to both Frege and Peano about it. It was, as Moore and Garciadiego comment, only when Russell received Frege's horrified reply that he realized how difficult the problem would be to solve. But the point I want to emphasize here is that, unlike Burali-Forti and Cantor, Russell realized from the first that he had a problem. The question for him was whether it was a big one or a small one; a trivial

[11] As distinct from Cantor's theological intent.

[12] We need to remember that, at this time, Russell was a radical realist who admitted many more things in the world, including sets, than were countenanced by most people's philosophies.

mistake in the reasoning or a matter that needed a fundamental reconstruction of logic. And the question for me, in the rest of this paper, is: Why was this so?

To seek the answer, we have to go back into Russell's philosophical past. Russell first discovered Cantor in 1896 and made at that time a detailed, but almost entirely critical, study of those of his works that Gösta Mittag-Leffler had put together in French translation in a special issue of his journal, *Acta Mathematica*, in 1883.[13] Russell returned to Cantor in 1899, 1900 and 1901 and no doubt subsequently, becoming steadily less critical in the process. But at the time of first encounter, Russell was a Hegelian struggling to give an account of continuous quantity.[14] Now Hegelians love contradictions and one might have expected to find Russell relishing the contradictions of continuous quantity. But Russell, like most of the late-nineteenth-century British Hegelians, was a consistentizing Hegelian. That is, he expected to find contradictions, but expected them to be progressively eliminated as one ascended through a hierarchy of levels of understanding towards an understanding of the Absolute. While Cantor associated the Absolute with his inconsistent multiplicities, Russell expected the Absolute to be entirely consistent, a realm in which the contradictions which infested lower levels (the realm of appearance rather than reality) would finally be resolved.[15]

Unusually for the British Hegelians, Russell's ascent to the Absolute was to be through a study of the special sciences, starting with geometry. Each special science, he thought, studied some aspect of appearance, but by welding them all together into a single philosophically well-thought out system he would (he hoped) approach a metaphysical science of the Absolute; a science of reality rather than of mere appearance. In accordance with this plan he set about collecting contradictions in the special sciences.[16] There was one particularly pervasive type, which he called the contradiction of relativity, which will not concern us here, although it was of great importance to Russell. It depended upon the doctrine of internal relations to be found in some of the British Hegelians, notably Bradley, and in 1898, in what amounted to a logical gestalt shift, he switched from regarding these contradictions as conclusions to be derived from a correct theory of relations, to seeing them as errors which invalidated the theory of relations which yielded them. It was this switch that ended his adherence to Hegelianism.

In his Hegelian efforts to deal with arithmetic and with continuous quantity one finds a predictable assemblage of traditional paradoxes about infinity. These were the sorts of paradoxes to be derived from Zeno and paradoxes to be found in the foundations of the calculus as to the impossibility of forming a finite quantity by summing infinitesimally small ones. Russell's consistentizing zeal was so great that,

[13] See ([23]: 97–8) for an account of the compilation, and ([*CPBR2*]: 460–81) for Russell's notes on it.

[14] See "On Some Difficulties of Continuous Quantity" [*CPBR2*]: 46–58.

[15] The metaphysical model for this was [2]. Russell's optimism that a theoretical articulation of the Absolute would be possible, however, comes from McTaggart.

[16] I have discussed this work at length in [24]. A much briefer summary can be found in [26].

for a time in 1895–6, he considered resolving these problems by a return to the pre-calculus tradition of finite indivisibles.[17]

There was one antinomy that Russell found in geometry which deserves our attention, and this is the antinomy of the necessary hypostatization of the point. During his Hegelian period Russell held a relational view of space and also the view he derived from Bradley that relations were nothing apart from their terms. Thus, on the one hand, geometry had nothing other than a system of relations to work with but, on the other, demanded terms for those relations. Points depended for their reality upon relations and relations for their reality upon points. In Russell's view, the geometrical point was conjured out of nothing to try (unsuccessfully) to resolve this contradiction.[18]

What I want to draw attention to in this is the criterion of reality that is implicit in it. For Russell, to be real was to be capable of being a term of a relation[19] and this, as he abandoned his Hegelianism and became a realist, became the principle that to be real was to be capable of being the subject of a proposition. This view he continued to hold—in a straight-forward, "what you see is what you get", kind of way—through *The Principles of Mathematics*. The doctrine persisted even beyond the *Principles*, but after 1905 his changing views on what was the logical subject of a proposition changed the import of the doctrine quite radically. It is this doctrine, I believe, that, at least partially, explains why Russell could not regard Cantor's inconsistent multiplicities, when he discovered them, with the same equanimity that Cantor did. Acknowledging such items at all required making them the logical subject of a proposition, and that, in turn, entailed that they were real, thus jeopardizing his search for a consistent account of reality.

Abandoning Hegelianism freed Russell at one stroke from many of the contradictions that had previously bothered him—all the contradictions of relativity were at once eliminated. Russell's delight in being free of them is evident in unpublished work he did in 1899, for example in "Fundamental Ideas and Axioms of Mathematics". But although the number of contradictions had been reduced, it had not been reduced to zero. "Fundamental Ideas" was to contain a chapter on the "Antinomy of Infinite Number". The chapter itself has not survived (if it was ever written) among the miscellaneous fragments of this book in the Russell Archives, but in his table of contents Russell summarizes the chapter as follows:

> *Antinomy of Infinite Number*. This arises most simply from applying the idea of totality to numbers. There is, and is not, a number of numbers. This

[17] He was influenced here by the French philosopher Arthur Hannequin, whose [27] he reviewed in *Mind* ([*CPBR2*]: 36–43).

[18] As a Hegelian Russell appealed to Boscovichian point atoms to provide something real to anchor geometrical points. After abandoning Hegelianism he adopted an absolute theory of space, cf. [*CPBR3*]: Part II.

[19] This should not be confused with his criterion for existence, which was that to exist is to have effects. In the technical terms of his (slightly) later philosophy, the criterion of reality could be stated as a criterion of being.

> and Causality[20] are the only antinomies known to me. This one is more all-pervading. It does *not* show finite numbers to be self-contradictory. Does it show the relation of part and whole to be so?[21] I think it does not. This relation seems to imply a totality, which is impossible, but on account of the terms, not the relation itself. No existing metaphysic avoids this antinomy. ([*CPBR2*]: 267)

The line of thought is remarkably similar to that which led Cantor to his inconsistent multiplicities, as is the following remark, also from his detailed table of contents, where he talks about "*all* concepts" and "*all* numbers":

> Totality here seems necessary; but if we make it so, infinite number with its contradictions becomes inevitable, being the number of concepts or of numbers. The only way to evade the contradiction is to deny the need of totality. ([*CPBR2*]: 266)

But Russell was still a long way from an understanding of Cantor's transfinite arithmetic. There is no trace, here, of Cantor's ascending sequence of alephs, nor, more importantly, of the crucial power-set theorem. Russell's understanding of the antinomy of infinite number is still distinctly pre-Cantorean. He explains the antinomy elsewhere in his table of contents:

> There are many numbers, therefore there is a number of numbers. If this be N, $N + 1$ is also a number, therefore there is no number of numbers. ([*CPBR2*]: 265)

But this, of course, is no paradox at all because the principle that $N + 1 > N$ on which it is based fails where N is an infinite cardinal.

It is not known exactly when Russell came to realize this. The 1899-1900 draft of *The Principles of Mathematics*, the draft he wrote before discovering Peano at the Paris Congress in early August 1900, reveals him, in many respects, not much further ahead.[22] He certainly takes Cantor more seriously here than previously. He considered Cantor's treatment of the continuum ([*CPBR3*]: 110–15), and a further chapter is devoted to transfinite numbers ([*CPBR3*]: 119–25). Cantor is credited with having created "the mathematical theory of infinity" and established "a branch of mathematics logically prior to the Calculus and even to irrationals". Nonetheless, Russell continues:

[20] Later in the table of contents, Russell has this to say about the antinomy of causality: "Each element has an effect, but no effect can be asserted apart from the whole. Illustration from the compounding of accelerations." ([*CPBR2*]: 271). Russell alludes to this problem in the Preface to [*POM*]: xvi–xvii.

[21] Russell was attempting at this time (before he discovered Peano) to found mathematics upon a fundamental part-whole relation. Husserl at the same time was working, quite independently, along similar lines.

[22] Despite having read [7] in July 1899. It was the *only* work he recorded having read that month ([*CPBR1*]: 362), which suggests he took especial care with it (or perhaps merely that he was too busy with other things). He re-read it in August 1900.

> I cannot persuade myself that his theory solves any of the philosophical difficulties of infinity, or renders the antinomy of infinite number one whit less formidable. Like most mathematical ideas on the subject, it consists of a skillful combination of the two sides of the antinomy in the proportions most useful for obtaining results. ([*CPBR3*]: 119)

He then presents a bunch of puzzles about the infinite, such as the traditional one that there are as many integers as even integers, and the new one that commutativity of addition fails for infinite ordinals ([*CPBR3*]: 121).[23] He does not, however, regard these points as in any way decisive against Cantor.

His chief objection remained his earlier antinomy of infinite number, but now recast in such a way as to avoid the mistaken assumption in the original version. Having read more Cantor by now, he is much more careful of the distinction between finite and infinite numbers. Consider the class of finite numbers and suppose it has a cardinal number n. Now either n is finite or infinite. If n is infinite it cannot be the number of finite numbers. But if n is finite there is a finite number $n+1$ greater than n. So, once again, n is not the number of finite numbers ([*CPBR3*]: 121–2).

Russell recognizes that Cantor's theory is designed to avoid this dilemma, but he works through a different version of the dilemma, this time starting, as it were, from above, with ω, and working down. Suppose the number of finite numbers is ω which is infinite, while "the number immediately preceding $\omega \ldots$ is finite". But if N is the number immediately preceding ω then N, not ω, is the number of finite numbers. To be the number of finite numbers, ω should be the last of the finite numbers, "not the number next after these" ([*CPBR3*]: 122).

He still misses the genuine paradox of the greatest cardinal, mainly because he remains reluctant to admit that Cantor's sequence of alephs (or omegas) can get off the ground.[24] "On the whole," he concludes, "it would seem impossible to maintain that any number is infinite" ([*CPBR3*]: 125). But, in coming to this conclusion, oddly enough, he takes a step towards the genuine paradox. The number of finite numbers, he argues, cannot be finite, because, for any finite number, adding 1 to it yields a greater number which is also finite. Therefore, either there is no number of finite numbers or, if there is, it is infinite. If, as Russell concludes, there are no infinite numbers, it follows that there is no number of finite numbers. This, in turn, conflicts with what he calls an axiom, namely the claim that "A given collection of many terms must contain some definite number of terms" ([*CPBR3*]: 124). That (finite) numbers form a genuine collection is shown, he argues, by the fact that "*all finite numbers*" is a legitimate concept; and he argues for this latter on the grounds that the concept can occur in a true proposition. The only way out, therefore, is to reject the axiom:

[23] It is a little puzzling why Russell should have regarded the fact that $\omega + 1 \neq 1 + \omega$ as a problem, since he was already aware of algebras where commutativity of addition fails from [33].

[24] Though he now knew much more about Cantor's work and even acknowledged its importance, this basic position was not very much different from the one he had adopted in 1896, cf. [*CPBR2*]: 51–2.

> Obvious as this axiom is, there is much to be said for rejecting it. Its acceptance, as we saw, leads to contradictions; its denial leads only to oddity. If we deny it, we may say, without actual contradiction, that there is no number of finite numbers ... In this case, a collection is infinite when portions of it consist of a number of terms, but the whole collection has no number; and the collection of numbers is infinite in this sense. ([*CPBR3*]: 125)

We don't know when this part of the 1899–1900 draft was written, but there is no doubt that Russell held to these conclusions right up until the Paris Congress. He comes to exactly the same conclusion in "Is Position in Time Absolute or Relative?" which he read at Oxford in May 1900 ([*CPBR3*]: 231). He made the same point in a footnote in his book on Leibniz:

> The general principle that all aggregates are phenomenal must not be confounded[25] with the principle, which Leibniz also held, that infinite aggregates have no number. This latter principle is perhaps one of the best ways of escaping from the antinomy of infinite number. ([*POL*]: 117n)

Interestingly, this footnote does not appear in the manuscript for the book and thus was added at a late stage in its preparation. The Preface for Russell's *Leibniz* is dated "September, 1900", i.e., *after* the Paris Congress, but it is known to be a very late addition,[26] so the best guess is that the footnote was added shortly before the Congress when Russell read proofs for the book.

Russell's attitude to the infinite in general and Cantor's work in particular changed dramatically immediately after he returned from the Paris Congress. In Russell's famous paper, "On the Logic of Relations", written in early October 1900 and published by Peano in 1901, transfinite numbers are defined and rules for the addition of infinite cardinals presented[27] apparently without any philosophical qualms. No doubt Russell's understanding of the technicalities of Cantor's transfinite arithmetic, and his ability to take the first steps towards logicising the theory, was helped by a second reading of [7] in August 1900. Moreover, a technical paper in symbolic logic like "The Logic of Relations" may not have seemed to Russell the best place to present his philosophical objections to the infinite. But these considerations can't explain the change in "The Logic of Relations", for the fact is, Russell's philosophical objections had vanished.

This is first made clear explicitly in his popular paper, "Recent Work on the Principles of Mathematics", written in January 1901, where he says that the problems of the

[25] One suspects from this that Russell himself had been guilty of confounding them.

[26] See [1]: vol. i, 9.

[27] Multiplication of infinite cardinals had to wait until [34]. Not all the advances of Russell's paper were made in October 1900, as Rodríguez-Consuegra ([30]: 157–62) points out—most notably Russell's famous logicist definition of the cardinal numbers was added later. For the October 1900 draft of the paper, see [*CPBR3*]: 590–612.

infinitesimal, the infinite, and continuity have been completely eliminated by Weierstrass, Dedekind and Cantor, moreover with solutions "so clear as to leave no longer the slightest doubt or difficulty" ([*CPBR3*]: 370)—a somewhat surprising judgment in view of the difficulty they had given Russell for the preceding four years. Even more ironical—in view of Russell's judgment, barely six months' earlier, that Cantor had merely combined "the two sides of the antinomy in the proportions most useful for obtaining results" ([*CPBR3*]: 119)—Russell now maintained that Cantor had "proceeded in the only proper way" which was to take "pairs of contradictory propositions, in which both sides of the contradiction would be usually regarded as demonstrable, and … strictly examine the supposed proofs" ([*CPBR3*]: 272–3). This procedure had revealed that the contradictions all depended upon a single principle namely, "if one collection is part of another, the one which is a part has fewer terms than the one of which it is a part and that, when once this maxim was rejected, all went well" ([*CPBR3*]: 373).

Russell, of course, was alive to the irony and was, indeed, exploiting it. He even borrowed some of his own language from the 1899-1900 draft of the *Principles*: "The Retention of this Axiom", he wrote in "Recent work" of the principle just stated about the relative cardinalities of collections where one is a proper part of the other, "leads to absolute contradictions, while its rejection leads only to oddities" ([*CPBR3*]: 375)—almost exactly quoting his earlier conclusion about the thesis that there is a number of numbers. "It was obvious", he now writes, "that there were infinities—for example, the number of numbers" ([*CPBR3*]: 372)—the negation of which he had carefully attempted to prove the previous year. And when he decried the efforts of previous philosophers as "unintelligible rigmarole" ([*CPBR3*]: 372), he certainly intended to include his own earlier work. This entire passage in "Recent Work" is to be seen as a private joke at the expense of Russell's work on infinity in the previous year.

What had happened to effect such a complete reversal? The answer is not just a more careful study of Cantor. That undoubtedly helped, but Russell had been studying Cantor off and on with some care since 1896, without thinking that the contradictions of infinity had been made one whit more tractable. The completeness of the reversal suggests another gestalt shift but, in this case, a more complicated one than that of converting *modus ponens* into *modus tollens*. The crucial difference, it seems to me, was that Russell, after the Paris Congress, had changed his logic. Prior to the Paris Congress Russell's logic had been built on an primitive part-whole relation. While this relation was at the heart of his logic, it seemed inescapable to sacrifice the 'obvious' truth that the number of numbers was infinite for the sake of preserving the principle that a proper part of a collection has fewer terms than the whole. Once he had embraced Peano's logic, the latter principle was no longer sacrosanct.

There is no doubt that Russell, in "Recent Work", thought that the problems of infinity had been completely and decisively solved. But pride comes before a fall, and Russell was to pay dearly for his hubris in "Recent Work". There is, indeed, one sign of trouble to come in the essay itself. Russell takes issue with Cantor's proof that there is no greatest number, in which he thinks that there is "some very subtle fallacy".

"[I]f this proof is valid", Russell says, "the contradictions of infinity would reappear in a sublimated form" ([*CPBR3*]: 375).

That this was the source of Russell's paradox is confirmed by what Russell said in *My Philosophical Development* almost sixty years later:

> I was led to this contradiction by considering Cantor's proof that there is no greatest cardinal number. I thought, in my innocence, that the number of all the things there are in the world must be the greatest possible number, and I applied his proof to this number to see what would happen . . . The application of Cantor's argument led me to consider the classes that are not members of themselves; and these, it seemed, must form a class. I asked myself whether this class is a member of itself or not. If it is a member of itself, it must possess the defining property of the class, which is to be not a member of itself. If it is not a member of itself, it must not possess the defining property of the class, and therefore must be a member of itself. ([*MPD*]: 75–6)

But the road to the paradox was far from straight-forward.

Russell had noticed the problem he refers to in "Recent Work" two months earlier and put it down to a mistake in Cantor's proof of the power-set theorem. He had written to Couturat about it on 8 December 1900:

> I have discovered a mistake in Cantor, who maintains that there is no largest cardinal number. But the number of classes is the largest number. The best of Cantor's proofs to the contrary can be found in [8]. In effect it amounts to showing that, if u is a class whose number is α, the number of classes included in u (which is 2^α) is larger than α. The proof presupposes that there are classes included in u which are not individuals [i.e. members] of u; but if $u = Class$, that is false: every class of classes is a class. ([31]: vol. i, 210–11).[28]

That this is the "very subtle fallacy" that Russell referred to in "Recent Work" the following month is made clear by his remarks on the topic in *The Principles of Mathematics* ([*POM*]: 362), to which I shall turn shortly.

Couturat replied that he didn't think it would be possible to admit the class of all classes without involving oneself in some kind of contradiction.[29] Russell, however, stuck to his guns:

> Concerning the class of classes, if you admit a contradiction in this concept, infinity will remain forever contradictory, and your works as well as Cantor's will not have resolved the philosophical problem. For there is a concept *class* and there are classes. Therefore, *class* is a class. Now it is proven (and this is essential to Cantor's theory) that all classes have

[28] The original letter is in French. The translation followed here is that provided at [*CPBR3*]: xxxii.

[29] Couturat to Russell, 3 January 1901 ([31]: vol. i, 218).

> a cardinal number. Thus, there is a number of classes, that is, a number of the class *class*. But it does not result in any contradiction, since the proof that Cantor gives that
>
> $$\alpha \in Nc. \supset 2^{\alpha} > \alpha$$
>
> presupposes that there is at least one class contained in a given class u (whose number is α) which is not itself a member of u, that is, one has:
>
> $$\exists \mathrm{Cls} \cap v \ni (v \supset . v \sim\in u). \qquad (1)$$
>
> If we put $u = \mathrm{Cls}$, this becomes false. Thus the proof doesn't hold.[30]

Russell, in fact, was mistaken in thinking that (1) was presupposed by Cantor's proof. It is not, however, clear when he realized this.

Interestingly, Couturat's letter of 3 January mentioned Burali-Forti's argument that the trichotomy law for infinite ordinals was false. "His reasoning", Couturat said, "was more specious than convincing" ([31]: vol. i, 218). Russell did not know of Burali-Forti's paper at time and asked Couturat for a copy of it. Nonetheless, he was inclined to agree with Burali-Forti's conclusion, and even to think the same thing held for cardinals—"Cantor's arguments on the subject are not conclusive" ([*SLBR1*]: 211; [31]: vol. i, 225). Here, too, Russell reversed himself the following year in "The General Theory of Well-Ordered Series" where he presented Cantor's proof of the trichotomy law for ordinals in Peanoesque guise ([*CPBR3*]: 389-421, *5.47). He dealt with Burali-Forti's paper in a note added in proof to the paper ([*CPBR3*]: 385)—which is, in fact, his first published response to any of the paradoxes—he accepted the trichotomy law, but still denied that there was a paradox because he now rejected the claim that the class of all ordinals can be well-ordered. There is nothing in this response which links Burali-Forti's result with his own paradox, which he had by then discovered. Indeed, Russell treats Burali-Forti's argument, as Burali-Forti himself had done, as a *reductio* argument.[31]

The precise route which led Russell from thinking that there was a subtle flaw in Cantor's argument that there was no greatest cardinal to the discovery of his paradox was first identified by Alberto Coffa, who found a passage in the manuscript for *The Principles of Mathematics* which was replaced in the published work and which makes Russell's thinking clearer.[32] The manuscript in question is that of Chapter XLIII, Part V of the *Principles* (Russell Archives file 230.030350-F14). It is part of what is usually known as "The Printers' Manuscript", but for reasons which will soon become

[30] Letter of 17 January, 1901, see ([*SLBR1*]: 211–12) and ([31]: vol. i, 225–6). We do not know whether "Recent Work" was written before or after this letter.

[31] Contrary to what is often suggested, he takes the same line in a slightly longer discussion in [*POM*]: 323. Garciadiego ([19]: 26–7) mistakenly cites this passage as the first statement of Burali-Forti's result as an "inconsistency in the theory of transfinite numbers"—as does Ferreirós; ([17]: 307).

[32] Coffa [14]. Prior to Coffa, Bunn ([3]: 239), Crossley [15], and Grattan-Guinness [22] had identified Russell's likely line of reasoning, and Grattan-Guinness was able to confirm it by reference to Russell's correspondence with G.H. Hardy and Philip Jourdain. Russell's letter to Hardy is reproduced in facsimile in ([19]: 196–201); the relevant parts of Russell's much less explicit letter to Jourdain are in ([21]: 52).

apparent, I shall refer to it more neutrally as "The F-14 Manuscript". The printed text begins to diverge from the manuscript at §344, which forms a little preamble to Russell's main argument.[33] The central concern, as is suggested by both Russell's later reminiscences and the contemporary correspondence with Couturat, is with the class of all classes or the class of all terms. In the manuscript (fol. 189), but not in the printed text, Russell gives an argument to show that the class of all classes can be put into one-one correspondence with the class of all terms,[34] since every class-concept is a term and every term a defines a class (viz., $\hat{x}(x = a)$). Moreover, since there is "a general proof, from the reflexiveness of similarity, that every class must have a number" (fol. 189), it follows that the class of terms and the class of classes have the same number, and, moreover, "since every other class is a proper part" of the class of all terms, it follows that this number must be the greatest possible. Thus, in the manuscript, the stage is set for a confrontation with Cantor's claim that there is no greatest number. The setting is much less elaborate in the published text: there he merely notes that there are certain classes—he mentions the class of all terms, the class of all classes and the class of all propositions—of which "it is easy to give an apparently valid proof that they have as many terms as possible". Thus, he concludes, "it would seem as though Cantor's proof [that there is no greatest cardinal] must contain some assumption which is not verified in the case of such classes" ([*POM*]: 362)—an appearance which in the published book, but not in the manuscript, he will go on to refute.

There were two arguments of Cantor's that he set out to rebut in the F-14 manuscript. The first ([7]: 44) was easily dealt with and Russell's treatment of it in the manuscript does not differ importantly from what was published in *Principles* (§345). The argument depended upon the assumption that the transfinite cardinals could be placed in one-one correspondence with the transfinite ordinals, and Russell simply rejected the assumption ([*POM*]: 363–4). He had earlier noted the lack of proof of the trichotomy law for cardinals and had expressed doubts whether any proof would be forthcoming ([*POM*]: 306).[35] Without the comparability of all cardinals, there was reason to doubt whether every set could be well-ordered. The set of all terms was a case in point. Russell saw no reason for thinking it could be well-ordered, but, if it could, then "the ordinals will have a perfectly definite maximum, namely that ordinal which represents the type of series formed by all terms without exception" ([*POM*]: 364). On the other hand, if the set of all terms cannot be well-ordered, "it is impossible to prove that there must be a maximum ordinal.[36] But in this case we may legitimately doubt whether there are as many ordinals as there are cardinals", and here he adverts once more to the failure of Cantor's proof [9] of the well-ordering of the cardinals. So far so good,

[33] A full list of variants between the F-14 manuscript and Part V of [*POM*] as published is given in [6].

[34] In the manuscript Russell uses 'individual' for what he elsewhere in the *Principles* calls 'terms'.

[35] He had assumed it as an axiom in a draft of "The Logic of Relations", cf. [*CPBR3*]: 596, *2·2) but did not assume it in the published version of the paper.

[36] In the published text, but not of course in the manuscript, Russell was able to cite Burali-Forti's argument [4], to which Couturat had drawn his attention, as further grounds for this conclusion.

and the F-14 manuscript for this section of the *Principles* (§345) corresponds closely to what was printed.

Cantor's second argument—his famous diagonal argument [8]—was a different matter entirely. Russell's published exposition of the proof ([*POM*]: §346) follows the F-14 manuscript (fols. 193–6) quite closely, as does the first paragraph of §347 outlining a simplified version of the proof. But the remainder of §347 and the next two sections of the published text do not appear in the manuscript. Instead, after stating Cantor's conclusion "that the number of classes contained in any class exceeds the number of terms belonging to [i.e., which are members of] the class" ([*POM*]: 366; fol. 196), Russell continues in the manuscript:

> Now if u be the class of classes, this is plainly self-contradictory, for classes contained in u will be only classes of classes, whereas terms belonging to u will be all classes without exception, so that the classes contained in u are a proper part of the class u itself. Hence there must be somewhere in Cantor's argument a concealed assumption not verified when u is the class of classes. (fol. 196)

Thus Russell's remark to Couturat, that Cantor tacitly assumes in his proof that there is at least one class contained in u which is not a member of u, was not based on Russell's having identified the point at which Cantor thus begs the question, but on Russell's assumption that he must have done so somewhere since the conclusion of his argument was plainly untenable.

But Russell goes beyond this in the manuscript. By applying Cantor's diagonal construction to *class*, the class of classes itself, he seeks to show that it does not produce a new class, a member of the power-set of *class* which is not in *class*. Cantor's proof of the power-set theorem proceeded by showing that for any function f from a class u to its power-set $\mathcal{P}(u)$, the class $w = \{x : x \in u \ \& \ x \notin f(x)\}$ is a member of $\mathcal{P}(u)$ but for all $y \in u$, $w \neq f(y)$. For suppose to the contrary that $y \in u \ \& \ w = f(y)$, $y \in w \leftrightarrow y \notin w$. Russell considers what would happen if we set $u = class$ and take f to be the function k from u to $\mathcal{P}(u)$ such that

$$k(x) = \begin{cases} \{x\}, & \text{if } x \text{ is a class but not a class of classes} \\ x, & \text{if } x \text{ is a class of classes} \end{cases}$$

In what follows Russell writes 'k_x' for the now-standard functional notation '$k(x)$':

> We have $u = class$, so that 'x is a u' means 'x is a class'. When x is not a class of classes, let k_x be the class of classes whose only member is x. When x is a class of classes, let k_x be x itself. Then we define a class u', in accordance with the above procedure, as containing every x which is not a member of its k_x, and no x which is a member of its k_x. Thus when x is not a class of classes, x is not a u'; when x is *class*, or *class of classes* or *class of classes of classes*, or *etc.*, x is not a u'; but when x is any other class of classes, x is a u'. Then Cantor infers

> that u' is not identical with k_x for any value of x. But u' is a class of classes, and is therefore identical with $k_{u'}$. Hence Cantor's method has not given a new term, and has therefore failed to give the requisite proof that there are numbers greater than that of classes. In fact, the procedure is, in this case, impossible; for if we apply it to u' itself, we find that u' is a $k_{u'}$, and therefore not a u'; but from the definition, u' should be a u'. In fact, when our original class consists of all possible combinations of all possible terms, the method, which assumes new combinations to be possible, necessarily fails, since, in this case, u' itself is a u. Thus what Cantor has proved is, that any power other than that of all classes can be exceeded, but there is no contradiction in the fact that this power cannot be exceeded. The exact assumption, in Cantor, which *class* fails to satisfy, is, that if u be the class whose power is to be exceeded, not all classes of u are themselves terms of u. (fols. 197–8)

It is evident now that Russell has come to the very brink of his paradox without realizing it, for his diagonal class u', is, a close companion of the class of all classes that are not members of themselves. It is defined by the condition

$$x \in u' \leftrightarrow x \notin k_x$$

so

$$u' \in u' \leftrightarrow u' \notin k_{u'}. \tag{2}$$

But u' is a class of classes, so by the definition of k,

$$u' = k_{u'}.$$

Substituting the identity in (2) gives

$$u' \in u' \leftrightarrow u' \notin u'.$$

Russell goes half-way to recognizing this in the passage just quoted, when he says that "the procedure ... is impossible; for if we apply it to u' itself, we find that u' is a $k_{u'}$, and therefore not a u'; but from the definition [of k], u' should be a u'". But, as Byrd ([6]: 66) points out, this line of argument is muddled up with another one, incompatible with it, which treats the construction of u' as legitimate, but then argues that, since u' is a class of classes, it is identical with $k_{u'}$ and so is not a "new term". On this approach, the diagonal construction produces a term, u', but not a new one; on the other approach, the construction is impossible because u' itself is inconsistent. Once the illegitimacy of u' is recognized, Russell's objection to Cantor collapses.

We know exactly when Russell wrote this remarkable passage, for the next folio of the manuscript is dated "Nov. 24, 1900". In retrospect, it seems odd that it took him another six months to discover the flaw in his reasoning. Between January and May 1901 we have little evidence of Russell's thinking on the paradoxes. The great exuberance of the period immediately following the discovery of Peano's work is over. During this period Russell found himself beset by personal problems (cf. [*SLBR1*]:

215–19), which no doubt took up much of his time. And no doubt he felt like a rest after his labours to complete the draft of the *Principles of Mathematics* by the end of 1900. In May, he started work on the book again, and it was in the course of this work that Russell discovered his paradox. The first extant statement of the paradox—though not, in fact, in its usual class form—occurs in Chapter III of the draft of Part I of *The Principles of Mathematics* that he wrote in May 1901.[37] What is striking about this first statement of the paradox is that it gives no hint of its Cantorean origins.

In the relevant chapter of the manuscript, Chapter III "Classes and Relations", Russell is thinking about homelier matters than these, about the need to be able to predicate concepts of themselves, as in "concept is conceptual", "term is a term", "1 is one" ([*CPBR3*]: 193–4). This is followed by a brief discussion of relations and their converses. Then Russell turns to "axioms relating classes and relations" and to the relation between predicates and classes, i.e., comprehension principles for classes though he doesn't call them that. He claims that "to have a given relation to a given term is a predicate" and that the terms having that relation to that term form a class. Moreover, "to have a given relation at all is a predicate, so that all references with respect to a given relation form a class" ([*CPBR3*]: 194). But then he notes that this last principle requires "some limitation . . . for the following reason":

> We saw that some predicates can be predicated of themselves. Consider now those . . . of which this is not the case. These are the referents (and also the relata) in a certain complex relation, namely the combination of non-predicability with identity. But there is no predicate which attaches to all of them and to no other terms. For this predicate will either be predicable or not predicable of itself. If it is predicable of itself, it is one of those references by relation to which it was defined, and therefore, in virtue of their definition, it is not predicable of itself. Conversely, if it is not predicable of itself, then again it is one of the said referents, of all of which (by hypotheses) it is predicable, and therefore again it is predicable of itself. This is a contradiction, which shows that all the referents considered have no common predicate, and therefore do not form a class. ([*CPBR3*]: 195)

This, of course, is not Russell's *class* paradox, but the predicational variant of it. It is perhaps not surprising, in view of Russell's logicist concerns, that he should have presented the paradox in this way—after all, predicates and relations were, from Russell's logicist point of view, more fundamental than classes. More surprisingly, in view of his later accounts of the discovery, he arrives at it not by any apparent considerations involving the class of all numbers, or the greatest cardinal, or the trichotomy law for ordinals, but by admitting that predicates may be predicated of themselves and then considering the relationship between predicates and classes. One could certainly come up with the paradox as Russell states it in the 1901 draft without knowing anything about Cantor.

[37] The manuscript is dated "May 1901" and thus rules out June as the month of the paradox's discovery.

It seems rather as if the paradox emerged from a consideration of the consequences of applying a predicate to itself. The problems deriving from the distinctness or identity of the predicate as predicated and the predicate as logical subject had plagued Russell's philosophy since he first embraced realism.[38] It seemed on the one hand that the two must be distinct—since one was a predicate and the other was a term—and yet they must be identical—since it was necessary on occasion to speak of the predicate and that required making it, and not another thing, the subject of a proposition. These problems remain in the *Principles* and even linger after the theory of ramified types in Russell's distinction between a relation as term and a relating-relation.

Indeed, in his first published presentation of the paradox in the *Principles* ([*POM*]: 88), it is linked to this very issue. The paradox of predication is said to emerge from treating "φ" as a separable part of "φx". It is rather, he asserts, that "φ" "lives in the propositions of the form φx, and cannot survive analysis".[39] But although he thinks such a view is necessary to avoid the paradox, he is not certain that it does not lead to contradictions of its own. He takes a similar sort of line when he comes to state the class form of the paradox in Chapter X of *Principles*. There he distinguishes between the class as one and the class as many, and links this to his embryonic type theory—the class as one is of the same logical type as its terms, and the class as many is not. It is, of course, significant that even in the first statement of the paradox he links the trouble (in effect) to unrestricted comprehension axioms for classes. For the predicational paradox itself this would hardly be necessary, but the motive behind his interest in predication was to provide a *logical* account of the notion of *class*. In the Preface to the *Principles* he admits that in this he has failed on account of the paradox ([*POM*]: xv–xvi).

What the predicational paradox shows, he claims in 1901, is "that not every definable collection of terms forms a class defined by a common predicate" or "A proposition containing one variable may not be equivalent to any proposition asserting that the variable in question belongs to a certain class" ([*CPBR3*]: 195). Even so, it is possible to make too much of this. In two major papers, "General Theory of Well-ordered Series" ([*CPBR3*]: 389–421) and the joint paper with Whitehead, "On Finite and Infinite Cardinal Numbers" ([*CPBR3*]: 425–30), written immediately after the May 1901 draft of the *Principles* Part I, he puts no restriction on the comprehension principle.[40]

In the 1901 draft of Part I, there is just one remark which suggests, albeit vaguely, a link between the predicational paradox and difficulties he had already been considering in connection with Cantor. In defining the class of predicates that are not predicable of themselves, he says,

[38] They are a special case of what I have elsewhere called "double aspect problems" [26]. They are endemic in Russell's philosophy.

[39] Interestingly, a view similar to this one was later taken up by the young Wittgenstein in 1912, who retailed it back to Russell as an alternative to his position then ([35]: 25).

[40] As late as 1905 in correspondence with Jourdain, Russell left the comprehension principle unrestricted, see [21]: 51–2.

> all those that are not predicable of themselves have been used up. The common predicate of all these predicates cannot be one of them, since for each of them there is at least one predicate (namely itself) of which it is not predicable. But again, the supposed common predicate cannot be any other predicate, for if it were, it would be predicable of itself. ([*CPBR3*]: 195)

The effect of this diagonalization argument is to bring the new paradox into the same form in which he had originally discussed the paradox of the number of numbers in the 1899–1900 draft of the *Principles*.

More surprising yet is the fact that in the 1901 draft of Part I there is little to suggest that Russell found the predicational paradox seriously disturbing. He does say that the need to restrict the comprehension principle "is of great importance in connection with the theory of the variable" ([*CPBR3*]: 195), but then he moves on. In view of all this it is tempting to suppose that when he included the predicational paradox in the May 1901 draft of Part I of the *Principles*, Russell *still* did not realize its connection to his earlier attempted refutation of Cantor's power-set theorem. This is not impossible, as noted already there is nothing in the predicational paradox that depends upon Cantor, and not even a thinker as quick as Russell can draw all the consequences of a problem at once—and for a writer as prolific as Russell a good deal of text can be produced in the interim. It may well have been that Russell's thinking about the relation between classes and predicates in the May 1901 draft led him to the predicational paradox independently of his criticism of Cantor's diagonalization argument, but this must then have led him very quickly to reconsider the legitimacy of the class u' with which he had challenged the power-set theorem the previous November. This could account for his uncertainty about the month in which the paradox was discovered. It may be, though this has to be sheer speculation, that he discovered the predicational paradox in May 1901, but it was not until the following month that he realized the flaw in his 'refutation' of Cantor and discovered the class form of the paradox.

The familiar class form of the paradox does not appear among Russell's surviving papers until some inconclusive working notes on the paradoxes written between March and April 1902 (now published as [*CPBR3*]: Appdx. II). Apart from these notes, we have a virtually complete silence from Russell on the paradox between its original discovery in May/June 1901 and the time he came to write the final version of Part I of *Principles*, almost exactly a year later. His correspondence with Couturat, for example, does not mention it. (Between May 1901 and May 1902 they corresponded mainly about Leibniz.) In almost the only reference to set-theoretic matters, Russell told Couturat that he now regarded Cantor's proof of the power-set theorem as "irrefutable".[41]

Even more surprising, is the fact that, when he completed the manuscript of the *Principles* in May 1902, he seems to have left his November 1900 critique of the power-set theorem unchanged. We know that Russell finished the book in a hurry and

[41] Letter of 2 October 1901 ([31]: vol. i, 259).

sent it off to the publisher as soon as it was done, writing to Cambridge University Press about it on the very day he finished work (cf. [*SLBR1*]: 236). Moreover, the published book shows signs of this haste and of the fact that the printers' manuscript was partly compiled from a variety of earlier drafts. Even so, it is difficult to believe that, a year after discovering the paradox and eight months after telling Couturat that he now thought Cantor's proof was irrefutable, Russell actually sent the F-14 manuscript to the printer. The grounds for doubt are not confined to the short section on Cantor. Byrd's collation of the F-14 manuscript with the published book reveals that none of the logicist definitions of Part V of ([*POM*]: §§253, 284, 295, 299) are in the manuscript. This confirms Rodríguez-Consuegra's ([30]: ch. 4) view that Russell did not embrace logicism immediately after the Paris Congress—indeed, the first draft of "The Logic of Relations" (Oct. 1900) and the F-14 manuscript (Nov. 1900) are non-logicist works. But this only makes our difficulty worse. Russell had been a logicist since at least January 1901; his May 1901 draft of Part I of the *Principles* defends logicism as does its May 1902 replacement. It seems barely credible that Russell should have packaged the latter along with the F-14 manuscript and send the result to his publisher when the two parts are in contradiction to each other on such important matters. It is one thing to suggest that he had forgotten what was in the F-14 manuscript—evidently he had, if he included it in the printer's manuscript—but it is another to suggest that he included it without even glancing at it, for even the most cursory inspection would have shown him how it varied from the position taken in those parts of the book that were written after January 1901.

Yet the alternative—that the F-14 manuscript, though early discarded, had somehow crept back into what is now known as "The Printer's Manuscript", replacing material written later—is not easy to support either. Outside the two areas of major variance between the F-14 manuscript and the published text of Part V, Byrd's collation reveals only relatively minor alterations. Russell rarely produced fair copy of a manuscript when isolated sheets from old ones could be cannibalized. So he would have been unlikely to produce a later document which was, for the most part, a fair copy of the F-14 manuscript. If he felt the need to revise a whole manuscript, he would typically rewrite the entire thing with only passing reference to the original, a process which resulted in many linguistic variants between the two. If he needed only to change certain parts of a paper, he would change only those parts, adding sheets and deleting passages as necessary, but leaving the other parts of the document in tact. Byrd's collation very much suggests the latter type of revision. Yet the F-14 manuscript contains no replacement sheets, no short folios, no longish cancelled passages, and, above all, no missing folios: there is, in short, no evidence that Russell ever made a major revision on this manuscript. The F-14 manuscript is plainly not what was left over after a subsequent (now lost) manuscript was prepared through a major round of revisions. The F-14 manuscript is in tact. If it was replaced before Russell sent *Principles* to the publisher, it was replaced in its entirety. But Russell rarely did this unless all parts of the manuscript required change. Byrd's collation strongly suggests that this was not the case with Part V of the *Principles of Mathematics*.

But if Russell did not change the F-14 manuscript before sending it to the printer, when did he change it? In proof? After he received Frege's response to the paradox? As early as October 1901, well before the printer's manuscript was prepared, he had admitted to Couturat that Cantor was right and he was wrong. He did not need to hear from Frege to know that the passage needed replacement.

The working notes of March/April 1902 are really the only evidence we have that Russell was looking for a solution in the right place. There the unrestricted comprehension principle for sets is clearly identified as the source of the class form of the paradox, cf. [*CPBR3*]: 563. Nonetheless, in the year following the discovery of the paradox, it is the predicational form of the paradox that seems to have been uppermost in Russell's mind. Not only does he introduce it separately from the other paradoxes in the *Principles*, but he gives it pride of place in Chapter X which is devoted to the paradoxes.[42] He presents it first and quite elaborately and then follows it with a brief restatement in terms of class-concepts and, finally, in terms of classes ([*POM*]: 102). It also gets pride of place in his famous letter to Frege of 16 June 1902, where the set-theoretic form is stated in Peano notation in a brief postscript ([32]: 124–5). It seems clear that, up to this point, Russell did not expect the problem to be hard to solve—certainly no harder than many of the other difficulties he had faced in the past four years. It seems clear, as he himself said long afterward, that he thought there was some mistake in his reasoning and one which, once identified, would not be difficult to correct. It also seems that, at the time he wrote to Frege, after a year of having tried and failed to find the mistake himself, he hoped that Frege would quickly point it out to him. No doubt he was almost as distressed at Frege's response as Frege had been to receive Russell's letter.

It is with the letter to Frege that the paradox begins its public history. So it is at this point that the real story begins. Nonetheless, since I promised only the prehistory of the paradox, it is at this point that I'll stop.

Acknowledgement. Research supported by the Social Sciences and Humanities Research Council of Canada. I am grateful to participants in the Russell 01-Conference, to Gregory H. Moore, and to an anonymous referee for comments on this paper.

Works by Russell

[*Auto*] 1969–1969. *The Autobiography of Bertrand Russell*, 3 vols. London: Allen and Unwin.

[*CPBR1*] 1983. *The Collected Papers of Bertrand Russell. Volume 1: Cambridge Essays, 1888–99*. Edited by K. Blackwell *et al*. London and Boston: Allen and Unwin.

[42] He gave it a chapter on its own in the plan for Part I of the *Principles* he drew up in April 1902—the month before he wrote the version that was actually published, cf. [*CPBR3*]: 211.

[*CPBR2*] 1990. *The Collected Papers of Bertrand Russell. Volume 2: Philosophical Papers, 1896–99*. Edited by N. Griffin and A. C. Lewis. London and Boston: Unwin and Hyman.

[*CPBR3*] 1993. *The Collected Papers of Bertrand Russell. Volume 3: Toward the* Principles of Mathematics*, 1900–02*. Edited by G. H. Moore. London and New York: Routledge.

[*CPBR11*] 1997. *The Collected Papers of Bertrand Russell. Volume 11: Last Philosophical Testament, 1943–68*. Edited by John G. Slater. London: Routledge.

[*MPD*] 1959. *My Philosophical Development*. London: Allen and Unwin.

[*POL*] 1900. *A Critical Exposition of the Philosophy of Leibniz*. Cambridge: Cambridge University Press. New ed., London: Allen and Unwin, 1937.

[*POM*] 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

[*SLBR1*] 1992. *The Selected Letters of Bertrand Russell. Volume 1: The Private Years, 1888–1914*. Edited by N. Griffin. London: Penguin.

Works by Other Authors

[1] Blackwell, Kenneth and Harry Ruja: 1994. *A Bibliography of Bertrand Russell*. London: Routledge.

[2] Bradley, Francis H.: 1893. *Appearance and Reality*. Oxford: Oxford University Press.

[3] Bunn, Robert: 1980. Developments in the foundations of mathematics, 1870–1910. In: I. Grattan-Guinness (ed.), *From the Calculus to Set Theory 1630–1910, An Introductory History*, London: Duckworth, 220–255.

[4] Burali-Forti, Cesare: 1897. A question on transfinite numbers. Translated by J. van Heijenoort. In [32]: 105–111.

[5] Burali-Forti, Cesare: 1897. On well-ordered classes. Translated by J. van Heijenoort. In [32]: 111–112.

[6] Byrd, Michael: 1994. Part V of *The Principles of Mathematics*. *Russell* n.s. 14: 47–86.

[7] Cantor, Georg: 1883. *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*. Leipzig: Teubner.

[8] Cantor, Georg: 1891. Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 1: 75–78.

[9] Cantor, Georg: 1895. Beiträge zur Begründung der transfiniten Mengenlehre (I). *Mathematische Annalen* 46: 481–512.

[10] Cantor, Georg: 1897. Beiträge zur Begründung der transfiniten Mengenlehre (II). *Mathematische Annalen* 49: 207–246.

[11] Cantor, Georg: 1899. Letter to Dedekind. Translated by S. Bauer-Mengelberg. In [32]: 113–117.

[12] Cantor, Georg: 1932. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Edited by E. Zermelo. Berlin: Springer.

[13] Cantor, Georg: 1991. *Briefe*. Edited by H. Meschkowski and W. Nilson. Berlin: Springer.

[14] Coffa, J. Alberto: 1979. The humble origins of Russell's paradox. *Russell* 33–4: 31–37.

[15] Crossley, John N.: 1973. A note on Cantor's theorem and Russell's paradox. *Australasian Journal of Philosophy* 51: 70–71.

[16] Dauben, Joseph: 1979. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Cambridge, MA: Harvard University Press.

[17] Ferreirós, José: 1999. *Labyrinth of Thought. A History of Set Theory and its Role in Modern Mathematics*. Basel: Birkhäuser.

[18] Frege, Gottlob: 1980. *Philosophical and Mathematical Correspondence*. Edited by G. Gabriel *et al*., abridged by Brian McGuinness, translated by Hans Kaal. Chicago: Chicago University Press.

[19] Garciadiego, Alejandro R.: 1992. *Bertrand Russell and the Origins of the Set-Theoretic "Paradoxes"*. Basel: Birkhäuser.

[20] Grattan-Guinness, Ivor: 1974. The rediscovery of the Cantor-Dedekind correspondence. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 76: 104–139.

[21] Grattan-Guinness, Ivor: 1977. *Dear Russell—Dear Jourdain. A Commentary on Russell's Logic Based on his Correspondence with Philip Jourdain*. London: Duckworth.

[22] Grattan-Guinness, Ivor: 1978. How Bertrand Russell discovered his paradox. *Historia Mathematica* 5: 127–137.

[23] Grattan-Guinness, Ivor: 2000. *The Search for Mathematical Roots. Logics, Set Theories and the Foundations of Mathematics from Cantor through Russell to Gödel*. Princeton: Princeton University Press.

[24] Griffin, Nicholas: 1990. *Russell's Idealist Apprenticeship*. Oxford: Clarendon Press.

[25] Griffin, Nicholas: 1991. Terms, relations, complexes. In: A. Irvine and G. Wedeking (eds.), *Russell and Analytic Philosophy*, Toronto: University of Toronto Press.

[26] Griffin, Nicholas: 2003. Russell's philosophical background. In: N. Griffin (ed.), *The Cambridge Companion to Russell*, Cambridge: Cambridge University Press.

[27] Hannequin, Arthur: 1895. *Essai critique sur l'hypothèse des atomes dans la science contemporaine*. Paris: Masson.

[28] Husserl, Edmund: 1994. *Early Writings on the Philosophy of Logic and Mathematics*. Dordrecht: Kluwer.

[29] Moore, Gregory H. and Alejandro Garciadiego: 1981. Burali-Forti's paradox: A reappraisal of its origins. *Historia Mathematica* 8: 319–350.

[30] Rodríguez-Consuegra, Francisco: 1991. *The Mathematical Philosophy of Bertrand Russell: Origins and Development*. Basel: Birkhäuser.

[31] Schmid, Anne-Françoise: 2001. *Bertrand Russell Correspondance sur la philosophie, la logique et la politique avec Louis Couturat (1897–1913)*, 2 vols. Paris: Editions Kimé.

[32] van Heijenoort, Jean (ed.): 1967. *From Frege to Gödel*. Cambridge, MA: Harvard University Press.

[33] Whitehead, Alfred N.: 1898. *Treatise on Universal Algebra*. Cambridge: Cambridge University Press.

[34] Whitehead, Alfred N.: 1902. On cardinal numbers. *American Journal of Mathematics* 24: 367–394.

[35] Wittgenstein, Ludwig: 1995. *Cambridge Letters. Correspondence with Russell, Keynes, Moore, Ramsey and Sraffa*. Edited by B. McGuinness and G. H. von Wright. Oxford: Blackwell.

[36] Zermelo, Ernst: 1908. A new proof of the possibility of a well-ordering. Translated by S. Bauer-Mengelberg in [32]: 183–198.

The Bertrand Russell Research Centre
McMaster University
1280 Main St. W.
Hamilton, Ontario
Canada L8S 4L6

E-mail: ngriffin@mcmaster.ca

Logicism's 'Insolubilia' and Their Solution by Russell's Substitutional Theory

Gregory Landini

Abstract. It is well known that Russell's 1903 *Principles of Mathematics* construed logic as a synthetic and a priori science of structure. Russell reified structures by adopting an ontology of propositions—mind and language independent 'states of affairs'. The thesis of Logicism advanced in the work held that the intuitions grounding all non-applied mathematics are logical intuitions of propositional structure. It has been widely reported that Logicism is dead and indeed that it died at Russell's own hands. Valiant as it was, the ramified type-theory of the 1910 *Principia Mathematica* did not save logicism from Russell's paradoxes (of classes and predication). The system offers no genuine 'solution' of the paradoxes, and requires an infinity axiom and an axiom of reducibility—neither of which can be counted among the truths of pure logic. Logicism is dead. Or is it? What has not been widely reported is that prior to composing the system of *Principia Mathematica*, Russell did solve the paradoxes. Applying his 1905 theory of incomplete symbols to form definite descriptions of propositions, he invented a type-free intensional calculus which emulates a simple type-theory of attributes (and thereby a simple type-theory of classes and relations in extension). This "no-classes" or "substitutional" theory, as it came to be called, has only recently begun to be investigated. This paper shows that Russell acted to hastily when he abandoned substitution in 1908. The substitutional theory may well resurrect logicism just as Russell had originally hoped.

1. Introduction

It is not well-known that prior to *Principia Mathematica* (1910) Russell formulated a theory that *solved* the paradoxes of classes and attributes plaguing logicism [10]. The theory was largely axiomatized by December of 1905 on the heels of Russell's theory of definite descriptions. But it was lost in unpublished manuscripts for almost seventy years. It is briefly discussed in Russell's 1905 paper "On Some Difficulties in the Theory of Transfinite Numbers and Order Types," and hinted at in the 1907 paper "Mathematical Logic as Based on the Theory of Types." But only one of Russell's detailed papers on the theory was published. Written first in English with the bold title "On 'Insolubilia' and Their Solution by Symbolic Logic," Russell translated the paper into French and published it in 1906 under the title "Les paradoxes de la logique." The paper was Russell's reply to Poincaré's paper "Les mathématiques et la logique"

which criticized logicism and offered the Vicious Circle Principle (*VCP*) as a solution of the paradoxes. Interest in the paper was confined to Russell's defense of logicism and his apparent agreement with Poincaré that violation of the *VCP* is the common source of both logical and semantic paradoxes. The hidden gems in Russell's paper, however, are his solution to the paradoxes and his *distinction* between syntactic and semantic paradoxes!

To solve the paradoxes, Russell formulated a substitutional theory of propositional structure. Russell's notion of substitution is unrelated to modern "substitutional quantification." Quantification in the theory is entirely objectual. There is only one style of variables in the calculus (entity/individual variables). Substitution is not a theory of types of entities, but rather a dissolution of the paradoxes by means of an eliminativistic ontological analysis and reconstruction of a type-stratified theory of attributes in intension (and thereby a type-stratified theory of classes and relations-in-extension). Unfortunately, Russell's eliminativistic methods have largely been unrecognized. In *Principia*, his thesis that classes are "logical constructions" has been interpreted to be ontologically reductive—classes are to be identified with certain entities called "propositional functions" (ramified and type-stratified attributes in intension), cardinal numbers are reduced to classes, ordinal numbers are reduced to (identified with) classes of well-ordering relations-in-extension, and so on. In his 1924 paper "Logical Atomism," Russell wrote

> One very important heuristic maxim which Dr. Whitehead and I found, by experience, to be applicable in mathematical logic, and have since applied in various other fields, is a form of Ockham's razor. When some set of supposed entities has neat logical properties, it turns out, in a great many instances, that the supposed entities can be replaced by purely logical structures without altering any of the detail of the body of propositions in question. This is an economy, because the entities with neat logical properties are always inferred, and if the propositions in which they occur can be interpreted without making this inference, the ground for the inference fails, and our body of propositions is secured against the need for a doubtful step. The principle may be stated in the form: 'Wherever possible, substitute constructions out of known entities for inferences to unknown entities.' ([26]: 326)

This statement of method is commonly interpreted as holding that inference to (the postulation of) new entities is to be replaced by logical constructions from entities with which we are acquainted such as sense-data and propositional functions. This is mistaken, however. The emergence of Russell's largely unpublished substitutional theory has revealed a completely different picture of Russell's conception of "logical construction." Russell's program is one of ontological eliminativism, reconceptualization, and structural realism. The ontology of an old theory is abandoned (or obviated), but structures of the old theory are re-conceptualized and recovered (where possible). The new theory *retains* the structures given by the laws of the old ontologi-

cal framework, just as Maxwell's equations for electromagnetic waves in an aether are retained in Einstein's no-aether theory of relativity. Russell's eliminativism is prominent in his thinking in the 20's and appears unequivocally in his neutral monism—a theory that attempts to reconstruct both the physics of continuants and the psychology of conscious states from an ontology of physical space-time events.[1] That Russell intended an eliminativistic approach is clear in his substitutional theory.[2] The major successes obtained by appeal to the existence of classes, the positive constructions of Cantor, Dedekind, Weierstrass, and Frege, are to be retained within substitution. But the results obtained by appeal to the existence of classes are conceptualized in an entirely new way. There is some loss, for instance, Cantor's transfinite ordinal number ω_ω cannot be recovered. But this loss is to be measured against the successes of the new theory. Indeed, had the theory yielded the conceptual successes that Russell had anticipated, one might venture to say that present mathematics would regard the notion of a class as present physics regards phlogiston, caloric fluid, and the aether.

It has been widely reported that logicism is dead, and indeed that it died at Russell's own hands. The doctrine of logicism espoused in Russell's *Principles* provided no solution of the paradoxes of classes and attributes. Valiant as it was, the ramified type-theory of *Principia* did not provide a solution either. The system relies upon an axiom (schema) of reducibility and an axiom of infinity, neither of which can be regarded as truths of pure logic. Logicism is dead. Or is it? Not all the options have been explored. Russell's substitutional theory has only recently been uncovered from his unpublished manuscripts. Collaborating with Whitehead, the substitutional theory was to have been set out in a companion volume to *Principles*. By 1905, Whitehead and Russell had decided to write a separate book *Principia Mathematica*, but Russell was convinced that *Principia* would have the substitutional theory at its foundation. In his paper "On the Substitutional Theory of Classes and Relations." Russell wrote that the theory "… affords what at least seems to be a complete solution of all the hoary difficulties about the one and the many; for while allowing that there are many entities, it adheres with drastic pedantry to the old maxim that 'whatever is, is one' " ([29]: 189). The paper was accepted for publication in October of 1906, but Russell withdrew it. In a letter to Jourdain he explained that "… there was much in it that wanted correction, and I preferred to wait till I got things into more finals shape. I am engaged at present in purging it of metaphysical elements as far as possible, with a view to getting the bare residuum on which its success depends."[3] Russell thought he had succeeded in this in his 1906 paper "On 'Insolubilia' and Their Solution by Symbolic Logic." As it heraldic title suggests, the paper promises a final solution of the paradoxes plaguing logicism. Alas, Russell abandoned substitution in 1907. Instead there was *Principia*, with legacy of interpreters working with an ontological reduction of classes to a ramified and

[1] In Russell's book *Philosophy* [27], we find an outright naturalization of epistemology and philosophy of mind.

[2] It is no less clear in *Principia*—if one knows were to look. *Principia* offers a no-propositions, no-classes, and *no-propositional functions* theory. But this is tangential to our present concerns, see [10].

[3] Quoted from [6]: 93.

type-stratified hierarchy of propositional functions. Russell's eliminativism, and the insights of the substitutional theory were lost. But we cannot rest with this as a proper end for the substitutional theory. I hope to show that the ideas Russell set forth for solving the 'insolubilia' can be employed to resurrect logicism.

2. Russell's *Distinction* between Logical and Semantic Paradoxes

By May of 1906 Poincaré's paper "Les mathématiques et la logique" appeared. Russell was eager to reply. Writing to Couturat on 15 May 1906, he remarked that "I still believe that my solution to the contradictions is good, but it seems to me that it needs to be extended to propositions. … I will follow your advice in replying to M. Poincaré. For this reason, I will not reply quickly, because I would like to get into order what I have to say about the solution of the contradictions."[4] Retreating to Covelly House for two months, Russell worked out a plan. By 14 June he wrote to Jourdain:

> The no-classes theory … shows that we can employ the symbol $\hat{y}\,\varphi y$ without ever assuming that this symbol in isolation means anything. I feel more and more certain that this theory is right. In order, however, to solve the *Epimenides*, it is necessary to extend it to *general* propositions, i.e., to such as $(x).\varphi x$ and $(\exists x).\varphi x$. This I shall explain in my answer to Poincaré's article. ([6]: 89)

At first blush, Russell's comments suggest that he was seeking a common solution to both the logical paradoxes and the semantic paradoxes. What actually transpired is quite different. The formal calculus of propositions underlying the substitutional theory has not been well understood. Studying Russell's brief comments in the 1907 paper "Mathematical Logic as Based on the Theory of Types," Quine once wrote that the central idea of substitution was that "instead of speaking of the class of all the objects that fulfill some given sentence, one might speak of the sentence itself and of substitutions within it. Now discourse about specified classes lends itself well enough to paraphrase in terms thus of sentences and substitution, but when we talk rather of classes in general, as values of quantifiable variables, it is not evident how to continue such paraphrase" [16]. Quine's difficulty in understanding how to continue the paraphrase stems from his eschewing the substitutional theory's ontology of propositions. The notion that a given entity is *in* (a constituent of) a proposition (state of affairs) is quite different than the notion of a singular term occurring in a sentence. During the era of substitution, Russell viewed propositions as mind and language independent states of affairs. Some propositions obtain, some do not. As such, propositions are intensional entities, not *intentional* entities. The existence of

[4] Letter to Couturat in [31]: 604.

propositions does not *ipso facto* introduce semantic notions into the object-language of the substitutional theory. Propositional versions of paradoxes such as the *Epimenides* are not germane to substitution. They can only appear in an applied form of a theory of propositions which introduces semantic concepts such as 'belief' and 'assertion' and regards these are direct relations between a mind and a proposition. This has not been appreciated. Indeed, Ramsey is often credited with having admonished Russell for thinking that both semantic paradoxes and the logical paradoxes of classes and attributes must have a common solution. In truth, Russell's substitutional theory did distinguish them and proposed quite distinct solutions.

Russell's manuscripts reveal that sometime after May of 1906 he had discovered a new paradox of propositions *unique* to substitution. I have called the new paradox Russell's "po/ao paradox." This paradox is not semantic, but logical, for its formulation does not depend upon introducing semantic notions such as "assertion," "belief," "designation," or "truth." Russell withdrew his May paper on the substitutional theory from publication because of the po/ao paradox of propositions, not because of the liar or any semantic paradox. He had decided not to respond too quickly to Poincaré because he needed to solve the po/ao paradox. The reason the po/ao paradox is not mentioned in Russell's reply to Poincaré is likely that Russell felt that its formulation is too technical and readers would not be familiar with details of the substitutional theory. The unfortunate consequence is that the fact that Russell's intention in 1906 to distinguish the logical and semantic paradoxes has been entirely lost.

A careful reading of Russell's reply to Poincaré reveals that he did *not* take the *VCP* to be the solution of the paradoxes. We find in ([29]: 206):

> It is important to observe that the vicious-circle principle is not itself the solution of vicious circle paradoxes, but merely the result which a theory must yield if it is to afford a solution of them. It is necessary, that is to say, to construct a theory of expressions containing apparent [bound] variables which will yield the vicious-circle principle as an outcome. It is for this reason that we need a reconstruction of logical first principles, and cannot rest content with the mere fact that paradoxes are due to vicious circles.

Russell maintains that the *VCP* is a regulative principle that any *solution*, grounded in logical re-constructions of first principles, must meet. The logical re-constructions Russell has in mind are important for they show that Russell distinguished *four* separate approaches to the paradoxes:

(1) A substitutional theory supplanting a type-theory of attributes in intension.

(2) A reconstruction of quantification theory without an ontology of general propositions.

(3) A recursive correspondence theory of "truth" for general statements.

(4) An hierarchy of languages.

In "On 'Insolubilia'"[5] and in "On the Substitutional Theory of Classes and Relations," the solution Russell offers for the Richard paradox, the König/Dixon paradox, and the Berry paradox is to adopt approach (4). On Russell's view, notions like "designation", "naming" are not univocal. One must first set out a formal language L, whose set of primitives and formation rules are determinate, and as a consequence "designation-in-L" and "namable-in-L" necessarily cannot be a primitives of L. Concerning the König/Dixon, Russell writes:

> Now the cardinal number of ordinals of the second class exceeds $\aleph_0$; hence some of these must be indefinable, and among those that are indefinable there must be a least. But this ordinal seems to be defined as 'the immediate successor of the ordinals that are definable.' At first this looks like a contradiction, but in fact it is not. For although every individual number less that this one is definable, the whole class of them is not definable. It *seems* to be defined as 'the class of definable ordinals'; but *definable* is relative to some given set of fundamental notions, and if we call this set of fundamental notions I, 'definable in terms of I' is never itself definable in terms of I. … It is easy to define 'definable in terms of I' by means of a larger apparatus I*; but then 'definable in terms of I*' will require a still larger apparatus I** for its definition, and so on. Or we may take 'definable in terms of I' as itself part of our apparatus, so that we shall now have an apparatus J consisting of I together with 'definable in terms of I.' In terms of this apparatus J, 'the least ordinal not definable in terms of I' is definable, but 'the least ordinal not definable in terms of J' is not definable. Thus the paradox of the least definable ordinal is only apparent. ([8]: 185)

This anticipates Tarski's hierarchy-of-language approach to semantic paradoxes. In contrast, Russell offers his approach (1) as a solution of the logical paradoxes of classes, attributes (propositional functions), Burali-Forti's contradiction of the greatest ordinal, and Cantor's contradiction of the greatest cardinal. These logical paradoxes are solved by the reconstruction of a type-theory of attributes in intension within the type-free substitutional "no-classes," "no-relations-in-extension," and "no-propositional functions" theory. The little-known po/ao paradox is solved by approach (2). In the early versions of the substitutional theory, it was assumed that any formula A could be nominalized to generate a singular term $\{A\}$. For example, a closed formula $(x)Bx$ involving bound variables could be nominalized to form $\{(x)Bx\}$ which is a singular term. In the intended interpretation $\{(x)Bx\}$ would be assigned a general proposition as its designation. Russell's later version of the substitutional theory, however, abandons the ontology of general propositions. Consequently, Russell had to reformulate quantification theory. In the new theory, a quantified formula $(x)Bx$ cannot be nominalized. Nonetheless, such a formula can flank the horseshoe sign because such

[5] See [29]: 209.

an occurrence is defined in terms of a prenex equivalent. For instance, where Bx is quantifer-free and x is not free in α Russell has:

$$(x)Bx \supset \alpha \stackrel{\text{def}}{=} (\exists x)(\{Bx\} \supset \alpha)$$

Quantification theory is reconstructed by a "no-general propositions" theory of logic in this manner. This is one of the "re-constructions of logical first principles" that Russell alludes to. Lastly, Russell offers approach (3) to generate a new account of truth and falsehood for general formulas. In the context of an ontology of propositions (states of affairs), 'truth' ('obtaining') and 'falsehood' ('non-obtaining') are taken to be primitive and unanalyzable properties. But when Russell abandoned his ontology of general propositions, a recursive correspondence theory of the truth-conditions of general statements is needed. The recursive theory generates a hierarchy of senses of "truth" and "falsehood" generated by the number of bound variables in the formulas they flank. This hierarchy is employed to solve the Liar paradox that arises with the statement "I am now asserting a false statement."

3. Logic as the Science of Propositional Structure

It is quite clear then, that Russell's awakening to the distinction between logical and semantic paradoxes did not arise with Ramsey's criticisms of *Principia*.[6] Peano had hinted that they should be distinguished in 1906 [13], and Russell explicitly offered different solutions for them in print. In 1905 Russell had considered belief as a dyadic relation to a proposition and investigated whether a solution of the *Epimenides Liar* (as a paradox of proposition rather than statements) would yield invaluable philosophical insights concerning the paradoxes of propositions. But he abandoned this view of belief when he abandoned general propositions. Semantic paradoxes are as tangential to the substitutional theory as they are to Zermelo set theory. Missing this important fact clouds one's understanding of the substitutional theory irremediably.

It is no less important to understand that the logical particles in the substitutional theory are not the logical particles of modern predicate calculi. In a modern predicate calculus, the logical particle "$\rightarrow$" is a statement connective; it is flanked by well-formed formulas (wffs) A and B of the language of the calculus to form a formula $A \rightarrow B$. Similarly, the modern logical particle "$\neg$" is a statement connective; it is flanked by a wff A to form a wff $\neg A$. In the language of substitution, the only variables are individual variables. Russell's logical particle "$\supset$" is a *dyadic predicate expression* for the relation of 'implication.' It is flanked by terms to form a formula. Accordingly, where α and β are any terms, $\alpha \supset \beta$ is a wff. (I use lower-case Greek for any singular term of the language of substitution.) The position of x and y in the formula $x \supset y$ are subject positions, and the individual (entity) variables x and y here are bindable,

[6] Unfortunately, Russell misleadingly attributes the distinction to Ramsey. See [28]: 126.

so that, $(x)(y)(x \supset y)$ is a formula of the language. (It says that for all x and y, x implies y.) Moreover, in the substitutional theory, a wff of the formal language can be nominalized to generate a genuine singular term. It is useful to use nominalizing braces "{" and "}" for nominalization. Thus, where x and y are individual variables, $\{x \supset y\}$ is a term. The distinction between terms and formulas is respected by Russell. The expression, $x \supset \{x \supset y\}$ is a formula. But since subject position is sufficient by itself to indicate a nominalizing transformation has taken place, we can often drop the brackets and use dots for punctuation. Thus instead of writing $x \supset \{x \supset y\}$ we can write, $x. \supset .x \supset y$ as our formula.[7]

To be sure, it may appear odd for entities such as people, rocks, and trees, to be said to stand in, or even fail to stand in, a relation of implication. Perhaps a relation $x|y$ of x's being non-copresent with y (adapted from the Peirce/Sheffer stroke)[8] is better suited for the role. But it is essential to a proper understanding of the substitutional theory *not* to identify its sign "$\supset$" with the modern statement connective "$\rightarrow$."

In an unpublished manuscript of 22 December 1905 entitled "On Substitution," Russell set out a formalization of the basic calculus for substitution [19]. From this we can reconstruct the formal calculus for the substitutional theory that Russell had in mind. In the original 1905 version, written prior to Russell's discovery of the po/ao paradox, the substitutional theory allowed any wff of its language to be nominalized to form a singular term. The substitutional language takes the following as primitive signs: (,), { , }, /, !, and $\supset$. The individual variables of the substitutional language are x, followed by one or more occurrences of "$\prime$". Informally we shall use any lower-case letter of the English alphabet. The terms are given inductively as follows. (1) All individual variables are terms; (2) If A is a wff, then $\{A\}$ is a term; (3) There are no other terms. The atomic wffs are of the form:

$$(x \supset y),$$
$$(x/y;\ z!u)$$

where x, y, z, u, are variables. The wffs are those of the smallest set K containing all atomic wff and such that $(\alpha \supset \beta)$, $(\alpha/\beta;\ \mu!\delta)$, and $(x)C$, and in K just when α, β, δ, μ, are any terms, C is any wff in K in which the individual variable x occurs free. The expression $\alpha/\beta;\ \mu!\delta$ says that δ results from substituting μ for every occurrence of β in α. The notion of *substitution* is primitive. Structurally the entity δ is exactly like α except that μ occurs *in* δ at exactly those positions (if any) at which β occurs *in* α.

The axiom schemata for the calculus for the original 1905 substitutional logic of propositions may be stated as follows:

$$\alpha. \supset .\beta \supset \alpha \qquad \text{S1}$$

$$\alpha. \supset .\beta \supset \delta :\supset: \beta. \supset .\alpha \supset \delta \qquad \text{S2}$$

[7] Russell himself took subject position to be sufficient to mark the nominalizing transformation and so did not employ brackets as we have above. In what follows I shall used dots symmetrically for punctuation.

[8] Of course, the Peirce/Sheffer stroke $A|B$ for alternative denial and $A \downarrow B$ for joint denial are statement connectives.

$\alpha \supset \beta :\supset: \beta \supset \delta. \supset .\alpha \supset \delta$ S3

$\alpha \supset \beta. \supset .\alpha :\supset: \alpha$ S4

$(u)Au \supset A[\alpha|u]$, where α is free for u in A. S5

$(u)(\alpha \supset Au). \supset .\alpha \supset (u)Au$, where u is not free in α. S6

$\alpha \ in \ \{A\alpha\}$ S7

$\alpha \ in \ \{A\beta_1, \ldots, \beta_n\} :\supset: \alpha = \{A\beta_1, \ldots, \beta_n\}.\vee.\alpha \ in \ \beta_1.\vee., \ldots, .\vee.\alpha \ in \ \beta_n$, where A is any wff all of whose distinct free terms are β_1,..., β_n. S8

$(x, y)(x \ in \ y \ .\&. \ y \ in \ x :\supset: x = y)$ S9

$(x, y, z)(x \ in \ y \ .\&. \ y \ in \ z :\supset: x \ in \ z)$ S10

$(p, a)(q)(x, y)(p/a; x!q \ .\&. \ p/a; y!q \ .\&. a \ in \ p :\supset: x = y)$ S11

$(p, a)(z)(q)(p/a; z!q \ .\&. \ a \ in \ p \ .\&. a \neq p :\supset: \ z \ in \ q \ .\&. z \neq q)$ S12

$(x, y)(x/x; y!y)$ S13

$(x, y)(x/y; y!x)$ S14

$(p, a)(x)(\exists q)(p/a; x!r. \equiv_r .q = r)$ S15

$(p)(\exists q)(q \ ex \ p)$ S16

$(\exists u)(\alpha \ out \ \{Au|v\}). \supset .(u)(\{A\alpha|v\}/\alpha; u!\{Au|v\})$, where α and u are free for v in A. S17

$(\exists u_1, \ldots, u_n)(\alpha \ out \ \{Au_1|v_1, \ldots, u_n|v_n\}) : \& : \alpha \neq \{Au_1|v_1, \ldots, u_n|v_n\}. \& .\alpha \neq \beta_1 . \& ., \ldots, . \& . \alpha \neq \beta_n. :\supset: . (x)(\exists u_1, \ldots, u_n)(\{A\sigma_1|v_1, \ldots, \sigma_n|v_n\}/\alpha; x!\{Au_1|v_1, \ldots, u_n|v_n\} : \& : \sigma_1|\alpha; x!u_1 . \& ., \ldots, . \& . \sigma_n|\alpha; x!u_n)$, where each u_i and σ_i, $1 \leq i \leq N$, are free for their respective v_i in A, and $\beta_1, \ldots, \beta_n$ are all the terms occurring free in A. S18

$\{(u)Au\} = \{(v)Av|u\}$, where v is free for u in A. S19

$\{(u)Au\} = \{(u)Bu\}. \supset .(u)(\{Au\} = \{Bu\})$ S20

$\{(u)Au\} \neq \{\alpha \supset \beta\}$ S21

$\alpha \ in \ \{(u)Au\} . \& . \alpha \neq \{(u)Au\} :\supset: (u)(\alpha \ in \ \{Au\})$, where α is not the individual variable u and u is not free in A. S22

$\{\alpha \supset \beta\} \neq\supset \ . \& . \{(u)A\} \neq\supset$ S23

The inference rules of the system are as follows:

Modus Ponens:

From A and $\{A\} \supset \{B\}$, infer B.

Universal Generalization:

From A infer $(u)A$,
where u is an individual variable free in A.

Replacement of Defined Signs:

Definiens and definiendum may replace one another in any context.

The following definitions are then introduced:

$$\sim\alpha \stackrel{\text{def}}{=} (x)(\alpha \supset x) \qquad \text{df}(\sim)$$

$$(\exists x)A \stackrel{\text{def}}{=} \sim(x) \sim A \qquad \text{df}(\exists)$$

$$\alpha \vee \beta \stackrel{\text{def}}{=} \sim\alpha \supset \beta \qquad \text{df}(\vee)$$

$$\alpha \,\&\, \beta \stackrel{\text{def}}{=} \sim(\alpha \supset \sim\beta) \qquad \text{df}(\&)$$

$$\alpha = \beta \stackrel{\text{def}}{=} (p)(a)(q)(r)(p/a; \alpha!q \,.\&.\, p/a; \beta!r :\supset: q \supset r) \qquad \text{df}(=)$$

$$(\exists x_1, \ldots, x_n)A \stackrel{\text{def}}{=} (\exists x_1), \ldots, (\exists x_n)A \qquad \text{df(E)}$$

$$(x_1, \ldots, x_n)A \stackrel{\text{def}}{=} (x_1), \ldots, (x_n)A \qquad \text{df(A)}$$

$$A \supset_{x_1,x_2,\ldots,x_n} B \stackrel{\text{def}}{=} (x_1)(A \supset_{x_2,\ldots,x_n} B) \qquad \text{df(I)}$$

$$\alpha \; out \; \beta \stackrel{\text{def}}{=} (u)(\beta/\alpha; u!\beta) \qquad \text{df(out)}$$

$$\alpha \; in \; \beta \stackrel{\text{def}}{=} \sim(\alpha \; out \; \beta) \qquad \text{df(in)}$$

$$\alpha \; ind \; \beta \stackrel{\text{def}}{=} \alpha \; out \; \beta \,.\&.\, \beta \; out \; \alpha \qquad \text{df(ind)}$$

$$\alpha \; ex \; \beta \stackrel{\text{def}}{=} \sim(\exists u)(u \; in \; \alpha \,.\&.\, u \; in \; \beta). \qquad \text{df(ex)}$$

These last definitions are worth discussing briefly. With the notion of substitution as a primitive, Russell defines what it is for one entity to be *out* (not a constituent) of another. Surely a is *out* of b if and only if every substitution of an entity u for a in b does not alter b. Russell then defines the notion of an entity being *in* (a constituent) of another as just the contradictory of *out*. On Russell's definition, there is a trivial sense in which every entity is *in* itself. But given Russell axioms (S9) and (S10), it follows that no entity is *in* itself in non-trivial sense. That is, no entity a is *in* an

another entity which is, in turn, *in a*. Russell also introduces the notion of an entity being *independent* of another: An entity *a* is independent of *b* if and only if neither is a constituent of the other. More important is Russell's definition of an entity's *excluding* another. An entity *a excludes b* if and only if no entity is a constituent of both *a* and *b*. This definition is important to axiom (S16) which is needed so that the antecedent clause of (S17) can be detached when needed. For if *a excludes b*, then since *a in a*, it follows that *a* is *out* of *b*. This reconstructs (with a bit of updating)[9] the system for the substitutional calculus of logic that Russell set out late in 1905.

4. Type-Theory as Formal Grammar

To understand how the substitutional language proxies a type-stratified second-order calculus with nominalized predicates, let us begin with a formal characterization of such a calculus. The primitive symbols of the language are (,), o, $'$, $\forall$, $\neg$, and $\rightarrow$. A *type symbol* is any expression that satisfies the following recursive definition:

1. "o" is a type symbol.

2. If $t_1, \ldots, t_n$ are all type symbols, then the expression $(t_1, \ldots, t_n)$ is also a type symbol.

3. These are the only type symbols.

The variables are small italic letters x with any type symbol superscript and followed by any number of occurrences of the sign "$'$". Informally, we shall use x^t, y^t, z^t, where t is a type symbol. When the type symbol t is not "o" we shall use Greek letters φ^t, θ^t, ψ^t, and also followed by one or more occurrences of "$'$" for convenience. The terms of the language are just the variables; the individual variables are those variables with type index "o", and the predicate variables are just the variables whose type index is not "o". The atomic formulae (wffs) of the language have the following form:

$$x^{(t_1,\ldots,t_n)}(y_1^{t^1}, \ldots, y_n^{t^n}). \tag{1}$$

The wffs of the language are those of the smallest set containing all atomic wffs and such that $(A \rightarrow B)$, $(\sim A)$, and $(\forall x^t)C$ are in K wherever A, B, and C are wffs in K and x^t is a variable. The axiom schemata for the type theory of attributes are then as

[9] The axiom schemata (S19), (S20), and (S21) are adapted from [1].

follows.

$$A. \rightarrow .B \rightarrow A \qquad \text{A1}$$

$$A. \rightarrow .B \rightarrow C :\rightarrow: A \rightarrow B. \rightarrow .A \rightarrow C \qquad \text{A2}$$

$$\sim B \rightarrow \sim A. \rightarrow .A \rightarrow B \qquad \text{A3}$$

$$(\forall x^t)(A \rightarrow B). \rightarrow .A \rightarrow (\forall x^t)B, \quad \text{where } x^t \text{ is not free in } A. \qquad \text{A4}$$

$$(\forall x^t)A \rightarrow A[y^t|x^t], \quad \text{where } y^t \text{ is free for } x^t \text{ in } A. \qquad \text{A5}$$

$$(\exists\varphi^{(t_1,\ldots,t_n)}(\forall x^{t_1},\ldots,x^{t_n})(\varphi^{(t_1,\ldots,t_n)}(x^{t_1},\ldots,x^{t_n}). \leftrightarrow .A), \quad \text{where } \varphi^{(t_1,\ldots,t_n)} \text{ does not occur free in } A. \qquad \text{A6}$$

Modus Ponens:

From A and $A \rightarrow B$, infer B.

Universal Generalization:

From A infer $(\forall u^t)A$, where u^t is in A.

Replacement of Defined Signs:

Definiens and definiendum may replace one another.

Definitions then include:

$$(\exists x^t)A \stackrel{\text{def}}{=} \neg(\forall x^t)\neg A$$

$$x^t = y^t \stackrel{\text{def}}{=} (\forall\varphi^{(t)})(\varphi^{(t)}(x^t). \leftrightarrow .\varphi^{(t)}(y^t))$$

$$A \wedge B \stackrel{\text{def}}{=} \neg(A \rightarrow \neg B)$$

$$A \vee B \stackrel{\text{def}}{=} \neg A \rightarrow B.$$

This completes the system.

Now to understand how the substitutional theory supplants a type-stratified theory of attributes, let us simply focus on the matter of comprehension principles. From (S17) and (S16) the following theorem schema is readily forthcoming:

$$(\exists p,\ a)(z)(a\ in\ p.\ \&\ .p/a;\ z!\{A\}), \quad \text{where } p \text{ and } a \text{ are not free in the wff } A. \qquad (\text{CP}_{\text{sub}})^1$$

From this, we get the following:

$$(\exists p,\ a)(z)(p/a;\ z. \equiv .\{A\}), \quad \text{where } p \text{ and } a \text{ are not free in the wff } A. \qquad (\text{CP})^1$$

The expression "$p/a; z$" is but a notational convenience. It is a definite description of an entity q just like p except for containing z wherever p contains a. Russell puts ([19]: 4):

$$p/a; z \stackrel{\text{def}}{=} (\iota q)(p/a; z!q).$$

Russell calls α/β of a definite description $\alpha/\beta;\delta$ its "matrix," and he speaks of the term α (and the proposition it names) as the "prototype" of the matrix. These notions, however, are apt to mislead. Russell speaks of the matrix as an "incomplete symbol" which has no meaning in isolation. But the incomplete symbol to focus upon is $(\iota q)(p/a; z!q)$, and this is simply a definite description. The theory of definite descriptions is straightforwardly applied to these new definite descriptions of propositions. Where $B(v)$ is a formula with v free, we have:

$$[(\iota q)(p/a; z!q)][B((\iota q)(p/a; z!q)|v)] \stackrel{\text{def}}{=} (\exists q)(p/a; z!r. \equiv_r .r = q : \& : B(q|v)).$$

All the same scope distinctions of definite descriptions, including the conventions on omission of scope marker when narrowest possible scope are intended, apply. It should be recalled that definite descriptions are not genuine terms of Russell's formal language. Accordingly, one cannot apply definitions which are framed in terms of genuine singular terms to definite descriptions. For example, in defining "&" Russell has:

$$\alpha \,\&\, \beta \stackrel{\text{def}}{=} \sim(\alpha \supset \sim \beta)$$

This cannot be applied to:

$$p/a; b \,\&\, s,$$

because α and β are expressions for genuine singular terms of the language. One must first eliminate the definite description, to yield:

$$(\exists q)(p/a; z!r. \equiv_r .r = q : \& : q \,\&\, s).$$

Then since q and s are variables one may apply df(&) to $q \,\&\, s$.

Russell goes on to define what he calls "simultaneous dual substitutions," "simultaneous triple substitutions," and so on. These are just carefully crafted successions of single substitutions. The expression "$s/t, w; r, c!q$," for instance, is for a dual substitution and says that q is exactly like s except containing r wherever s contains t and containing c wherever s contains w. One must take care in the definition so that q is the entity intended—since, for example, the substitution of r for t may remove w from s. Russell's definitions are complicated and we shall avoid discussion of them here, for there is a simpler approach. There is no need to define "simultaneous substitutions." We can simply put:

$$p/a, b; x, y!q \stackrel{\text{def}}{=} (\exists e,\ h,\ t)(p/a; e!h : \& : h/b; y!t. \,\&\, .t/e; x!q). \qquad \text{df(dual)}$$

For any formula $A(u,\ v)$, we can employ (S16) to find appropriate entities p, a, and b, so that

$$(x,\ y)(p/a, b; x, y!\{A(x|u, y|v)\}).$$

Accordingly, we arrive at the following theorem schema for dual substitutions:

$$(\exists p,\ a,\ b)(x,\ y)(p/a, b; x, y. \equiv .\{A\}), \qquad (\mathrm{CP})^2$$

where p and a and b are not free in the wff A. Here the expression $p/a, b; x, y$ is a convenient way of writing the definite description, $(\iota q)(p/a, b; x, y!q)$. In a similar way, we can go on to triple substitutions and the theorem schema:

$$(\exists j,\ k,\ u,\ v)(q,\ p,\ a)(j/k, u, v;\ q, p, a. \equiv .\{A\}), \qquad (\mathrm{CP})^3$$

where j and k and u and v, are not free in the wff A. The process continues for any finite number of substitutions.

The comprehension theorem schemata of substitution recover instances of the comprehension axiom schema (A6) of simple second-order type theory. Consider the following pairs:

$$(\exists\varphi^{(\mathrm{o})})(\forall x^{\mathrm{o}})(\varphi^{(\mathrm{o})}(x^{\mathrm{o}}). \leftrightarrow .x^{\mathrm{o}} = x^{\mathrm{o}}) \qquad \#1$$

$$(\exists p,\ a)(x)(p/a; x. \equiv .x = x) \qquad \#1\mathrm{s}$$

$$(\exists\varphi^{((\mathrm{o}))})(\forall\psi^{(\mathrm{o})})(\varphi^{((\mathrm{o}))}(\psi^{(\mathrm{o})}). \leftrightarrow .(\forall x^{\mathrm{o}})\neg\psi^{(\mathrm{o})}(x^{\mathrm{o}})) \qquad \#2$$

$$(\exists q,\ p,\ a)(r,\ c)(q/p, a; r, c. \equiv .(x) \sim (p/a; x)) \qquad \#2\mathrm{s}$$

$$(\exists\varphi^{(\mathrm{o},\mathrm{o})})(\forall x^{\mathrm{o}})(\forall y^{\mathrm{o}})(\varphi^{(\mathrm{o},\mathrm{o})}(x^{\mathrm{o}}, y^{\mathrm{o}}). \leftrightarrow .x^{\mathrm{o}} = y^{\mathrm{o}}) \qquad \#3$$

$$(\exists q,\ p,\ a)(r,\ c)(q/p, a; r, c. \equiv .r = c) \qquad \#3\mathrm{s}$$

The first of the pairs are instances of the comprehension axiom schema (A6); the second are their translations into the language of the substitutional theory. The notion of type has been built into the logical grammar of substitution. It will be noted that #2s and #3s both employ dual substitutions, yet #2s proxies a type ((o)) attribute (an attribute of attributes of individuals) and #3s proxies a type (o,o) relation between individuals. This shows that the notion of type does not correspond simply to the number of substitutions employed in the substitutional theory. The structure is central as well. Nonetheless, it is clear that monadic predications are captured in the substitutional theory by using expressions such as "$p/a; x!t$" and "$q/p, a; r, c!t$" and so on. Accordingly, the expression "$\varphi(\varphi)$" cannot be proxied. It would require a sequence such as "$p/a;\ p/a!t$" and this, as well as its negation, is ungrammatical.

The substitutional theory is entirely type-free. It countenances no types of entities, and there is only one style of variables—individual variables. (Recall that every lowercase letter of the English alphabet is used as an individual variable for convenience. There are no special "propositional" variables in the theory.) The type regimented grammar of simple type theory is built into the substitutional language's replacement of the use of predicate variables. At the same time, it should be clear enough from the

above that any instance of the comprehension schema A6 in the primitive notation of second-order type theory is translatable into the language of substitution. Indeed, any formula in primitive notation of simple type theory can be translated into the language of substitution. Types become part of logical grammar.

5. Classes in Substitution

Class symbols are incomplete symbols in the substitutional theory and their use is much as we find in *Principia Mathematica*. The class symbol $\hat{y}^{o}(Ay^{o})$ is supplanted by use of

$$\iota(p/a)[p/a;\ y \equiv_y Ay].$$

For example, to reconstruct the theorem,

$$(\forall x^{o})(x^{o} \in \hat{y}^{o}(y^{o} = x^{o})).$$

Russell puts:

$$(x)(x \in \iota(p/a)[p/a\ .\ \equiv_y\ .\ y = x]).$$

The use of such class symbols in substitution is supported by definitions governing each context of their use. For instance, there are the following:

$$z \in \iota(p/a)[p/a; x \equiv_x Ax] \stackrel{\text{def}}{=} (\exists p,\ a)(p/a; x \equiv_x Ax.\,\&\,.z \in p/a)$$

$$x \in p/a \stackrel{\text{def}}{=} (\exists q)(p/a; x!q.\,\&\,.q).$$

$$\iota(p/a)[p/a; x \equiv_x Ax] = \iota(p/a)[p/a; x \equiv_x Bx] \stackrel{\text{def}}{=}$$
$$(\exists p,\ a)(p/a; x \equiv_x Ax.\,\&\,.(\exists l,\ m)(l/m; x \equiv_x Bx.\,\&\,.p/a = l/m))$$

$$p/a = l/m \stackrel{\text{def}}{=} p = l.\,\&\,.a = m$$

Similarly, in the next type, a class $\hat{y}^{(o)}(By^{(o)})$ of type ((o)) is supplanted by use of :

$$\iota(s/t, w)[s/t, w.\ \approx_{r,c}\ .B(r,\ c)].$$

To take a simple illustration, define as follows:

$$\Lambda x \stackrel{\text{def}}{=} x \neq x.$$

$$0(r,\ c) \stackrel{\text{def}}{=} (z) \sim (z \in \iota(p/a)[p/a; x \equiv_x r/c; x]).$$

The substitutional reconstruction of $\Lambda^{(o)}\varepsilon 0^{((o))}$, becomes:

$$\iota(p/a)[p/a; x \equiv_x \Lambda x] \in \iota(s/t, w)[s/t, w \approx_{r,c} 0(r,\ c)].$$

This is supported by the following definitions:

$$\iota(p/a)[p/a; x \equiv_x Ax] \in \iota(s/t, w)[s/t, w \approx_{r,c} B(r, c)] \stackrel{\text{def}}{=} (\exists s, t, w)(s/t, w \approx_{r,c} B(r, c). \&.(\exists p, a)(p/a; x \equiv_x Ax. \&.p/a \in s/t, w))$$

$$s/t, w \approx_{r,c} B(r, c) \stackrel{\text{def}}{=} (\exists l, m)(l/m; x \equiv_x r/c; x. \&.l/m \in s/t, w) \equiv_{r,c} (\exists l, m)(l/m; x \equiv_x r/c; x. \&.B(l, m)).$$

$$\iota(s/t, w)[s/t, w \approx_{r,c} A(r, c)] = \iota(s/t, w)[s/t, w \approx_{r,c} B(r, c)] \stackrel{\text{def}}{=} (\exists s, t, w)(s/t, w \approx_{r,c} A(r, c). \&.(\exists h, d, e)(h/d, e \approx_{r,c} B(r, c). \&.s/t, w = h/d, e))$$

$$s/t, w = h/d, e \stackrel{\text{def}}{=} s = h. \&.t = d. \&.w = e.$$

Type-stratified class abstraction

$$z^o \in \hat{y}^{\text{o}}(Ay^{\text{o}}) \leftrightarrow Az^{\text{o}}$$

is now supplanted by:

$$z \in \iota(p/a)[p/a; x \equiv_x Ax]. \equiv .Az.$$

In the next type, class abstraction,

$$\hat{y}^{\text{o}}(Ay^{\text{o}}) \in \hat{y}^{(\text{o})}(By^{(\text{o})}) \leftrightarrow B(\hat{y}^{\text{o}}(Ay^{\text{o}}))$$

is supplanted by:

$$\iota(p/a)[p/a; x \equiv_x Ax] \in \iota(s/t, w)[s/t, w \approx_{r,c} B(r, c)]. \equiv . (\exists p, a)(p/a; x \equiv_x Ax. \&.B(p, a)).$$

Working through the definitions, it is clear that class abstraction schemata for each type are forthcoming as theorem schemata from the existence theorems afforded by $(\text{CP})^1$ and $(\text{CP})^2$. Similar constructions and definitions permit the generation of a substitutional proxy for the type-stratified theory of relations-in-extension.

6. The po/ao Paradox and Russell's 'Insolubilia'

The 1905 substitutional theory is inconsistent. By April of 1906, Russell discovered his po/ao paradox.[10] Its existence as a new paradox and its unique significance for the historical development of Russell's ramified type-theory were largely unknown until I unearthed it from the archival manuscripts.[11] The problematic axiom schema is (S17).

[10]The paradox is the central theme of a number of Russell's worknotes for 1906. See ([22]: 7, 57, 71) and [24].

Coupled with (S16) one arrives at the theorem schema $(\mathrm{CP}_{\mathrm{sub}})^1$. The paradox[12] is derived from the following instance:

$$(\exists t,\ w)(x)(t/w;\ x!\{x = \{p \supset a\}.\ \&\ .p/a;\ x!r :\supset_{p,a,r}: \sim r\}).$$

By existential instantiation we arrive at:

$$(x)(t/w;\ x!\{x = \{p \supset a\}.\ \&\ .p/a;\ x!r :\supset_{p,a,r}: \sim r\}).$$

Then by universal instantiation we have:

$$t/w;\ \{t \supset w\}!\{\{t \supset w\} = \{p \supset q\}.\ \&\ .p/a;\ \{t \supset w\}!r :\supset_{p,a,r}: \sim r\}.$$

It is provable in substitution that:

$$(p,\ a)(r,\ c)(\{p \supset a\} = \{r \supset c\} :\supset:\ p = r.\ \&\ .a = c).$$

We arrive at the following contradiction:

$$\begin{aligned}\{t \supset w\} =& \{p \supset a\}.\ \&\ .p/a;\ \{t \supset w\}!r :\supset_{p,a,r}: \sim r\\ & .:\equiv:.\sim(\{t \supset w\} = \{p \supset a\}.\ \&\ .p/a;\ \{t \supset w\}!r :\supset_{p,a,r}: \sim r).\end{aligned}$$

The flaw in the substitutional system lies with schema (S17).

But what precisely is wrong with (S17)? Russell's assessment in May of 1906 was that the source of the contradiction lies in his assumption of general propositions. In his article "On 'Insolubilia' and Their Solution by Symbolic Logic," Russell abandoned his ontological commitment to general propositions. Of course, the abandonment of general propositions should not be confused with the abandonment of quantified formulas.[13] The employment of quantified wffs does not make any commitment to general propositions (states of affairs). But the affect of abandoning general propositions is that it is no longer possible to nominalize a wff $(x)Ax$, to form a term of the form $\{(x)Ax\}$. Only quantifier-free formulas can be nominalized. This is significant because Russell's sign "$\supset$" is a dyadic predicate constant that must be flanked by terms. In the new theory,

$$\{(x)(x = x)\}.\ \supset .q \supset \{(x)(x = x)\},$$

is not a proper instance of axiom schema (S1). Quantification theory has to be modified to accommodate the abandonment of general propositions. Russell therefore reformulates quantification theory by defining subordinate occurrences of quantified

[11] Cocchiarella showed the way. He observed that Russell had blundered in thinking that substitution would be able to capture the Cantorian fact that there must be more classes of propositions than propositions. See [2].

[12] Russell's original formulation proceeded as follows. Abbreviate putting:

$p_{\mathrm{o}} \overset{\mathrm{def}}{=} \{a_{\mathrm{o}} = \{p/a;\ b!q\}.\ \&\ .p/a;\ a_{\mathrm{o}}!r :\supset_{p,a,r}: \sim r\}.$

Then observe that:

$p_{\mathrm{o}}/a_{\mathrm{o}};\ \{p_{\mathrm{o}}/a_{\mathrm{o}};\ b!q\}!\{\{p_{\mathrm{o}}/a_{\mathrm{o}};\ b!q\} = \{p/a;\ b!q\}.\ \&\ .p/a;\ \{p_{\mathrm{o}}/a_{\mathrm{o}};\ b!q\}!r :\supset_{p,a,r}: \sim r\}$

A contradiction follows since the following is a theorem,

$(p, a)(r, c)(b, q)(\{p/a;\ b!q\} = \{r/c;\ b!q\} :\supset:\ p = r.\ \&\ .a = c).$

[13] Surprisingly, some commentators are guilty of just this confusion. See [15].

formulas in terms of an equivalent formula in prenex normal form.[14] The full revision of quantification theory is not set out in "On 'Insolubilia'," but it does not impose insurmountable problems. Indeed, a system very like it appeared in *9 of the 1910 *Principia*—though it plays a quite different role in that work.

The basic system Russell had in mind for "On 'Insolubilia'" is then the following. The primitive signs for the substitutional language takes as primitive signs: (,), {, }, /, !, $\supset$, **f**, and $\exists$. The individual variables of the substitutional language are x, followed by one or more occurrences of "′". Informally we shall use any lower-case letter of the English alphabet. The terms are given inductively as follows. (1) All individual variables are terms; (2) If A is a quantifier-free wff, then $\{A\}$ is a term; (4) There are no other terms. The atomic wffs are: **f**, $(x \supset y)$, and $(x/y; z!u)$, where *x, y, z, u* are variables. The wffs are those of the smallest set K containing all atomic wff and such that $(\alpha \supset \beta)$, $(\alpha/\beta; \mu!\delta)$, and $(x)C$ and $(\exists x)C$ are in K just when α, β, δ, μ, are any terms, and C is any wff in prenex-normal form in K in which the individual variable x occurs free. The axiom schemata for the calculus are as follows:

$\alpha. \supset .\beta \supset \alpha$ S1

$\alpha. \supset .\beta \supset \delta :\supset: \beta. \supset .\alpha \supset \delta$ S2

$\sim\sim\alpha \supset \alpha$ S3

$\alpha = \beta. \supset .A[\alpha|v,\ \alpha] \supset A[\beta|v,\ \alpha]$, where α, β are free for v in A. S4

$A[\alpha|u] \supset (\exists u)Au$, where α is free for u in A. S5

$A[\alpha|u] \vee A[\beta|u] \supset (\exists u)Au$, where α, β are free for u in A. S6

$\alpha\ in\ \{A\alpha\}$ S7

$\alpha\ in\ \{A\beta_1, \ldots, \beta_n\} :\supset: \alpha = \{A\beta_1, \ldots, \beta_n\}.\vee.\alpha\ in\ \beta_1.\vee., \ldots, .\vee.\alpha\ in\ \beta_n$, where A is any wff all of whose distinct free terms are β_1,..., β_n. S8

$(x, y)(x\ in\ y\ .\&.\ y\ in\ x :\supset: x = y)$ S9

$(x, y, z)(x\ in\ y\ .\&.\ y\ in\ z :\supset: x\ in\ z)$ S10

$(p, a)(q)(x, y)(p/a; x!q\ .\&.\ p/a; y!q\ .\&.\ a\ in\ p :\supset: x = y)$ S11

$(p, a)(z)(q)(p/a; z!q\ .\&.\ a\ in\ p\ .\&.\ a \neq p :\supset: z\ in\ q\ .\&.\ z \neq q)$ S12

$(x, y)(x/x; y!y)$ S13

$(x, y)(x/y; y!x)$ S14

$(p, a)(x)(\exists q)(p/a; x!r. \equiv_r .q = r)$ S15

$(p)(\exists q)(q\ ex\ p)$ S16

[14] In *Principia Mathematica's* *9, we get a glimpse of how the calculus would be formulated. *Principia*, however, regards its logical particles *statement* connectives because it explicitly abandons the ontology of propositions.

$(\exists u)(\alpha\ out\ \{Au|v\}).\supset .(u)(\{A\alpha|v\}/\alpha; u!\{Au|v\})$, S17
where α and u are free for v in A.

$(\exists u_1, \ldots, u_n)(\alpha\ out\ \{Au_1|v_1, \ldots, u_n|v_n\}) : \& : \alpha \neq$ S18
$\{Au_1|v_1, \ldots, u_n|v_n\}. \& .\alpha \neq \beta_1 . \& ., \ldots, . \& . \alpha \neq \beta_n. :\supset: .$
$(x)(\exists u_1, \ldots, u_n)(\{A\sigma_1|v_1, \ldots, \sigma_n|v_n\}/\alpha; x!\{Au_1|v_1, \ldots, u_n|v_n\}$
$: \& : \sigma_1|\alpha; x!u_1 . \& ., \ldots, . \& . \sigma_n|\alpha; x!u_n)$,
where each u_i and σ_i, $1 \leq i \leq N$, are free for their respective v_i
in A, and $\beta_1, \ldots, \beta_n$ are all the terms occurring free in A.

The inference rules of the system are as follows:

Modus Ponens$_1$:

From A and $\{A\} \supset \{B\}$, infer B.

Modus Ponens$_2$:

From A and $A \supset B$, infer B.

Universal Generalization:

From A infer $(u)A$,
where u is an individual variable free in A.

Switch

From $B[(u)(\exists v)A]$, infer $B[(\exists v)(u)A]$,
where all free occurrences of the variable u in A are on one side of
a logical particle and all free occurrences of the variable v
in A are on the other.

Replacement of Defined Signs:

Definiens and definiendum may replace one another in any context.

Where α and β are any quantifier-free terms of the language, and A and B are any formulas of the language, the following are definitions:

$$x = y \stackrel{\text{def}}{=} x/x; y!x \qquad \text{df}(=)$$

$$\sim\alpha \stackrel{\text{def}}{=} \alpha \supset \mathbf{f} \qquad \text{df}(\sim)_1$$

$$\sim(u)A \stackrel{\text{def}}{=} (\exists u) \sim A \qquad \text{dfs}(\sim)_2$$

$$\sim(\exists u)A \stackrel{\text{def}}{=} (u) \sim A \qquad \text{dfs}(\sim)_3$$

$$\alpha \vee \beta \stackrel{\text{def}}{=} \sim\alpha \supset \beta \qquad \text{df}(\vee)_1$$

$$A \vee B \stackrel{\text{def}}{=} \sim A \supset B \qquad \text{df}(\vee)_2$$

$$\alpha \,\&\, \beta \stackrel{\text{def}}{=} \sim(\alpha \supset \sim \beta) \qquad \text{df}(\&)_1$$

$$A \,\&\, B \stackrel{\text{def}}{=} \sim(A \supset \sim B) \qquad \text{df}(\&)_2$$

$$\alpha \equiv \beta \stackrel{\text{def}}{=} (\alpha \supset \beta)\,\&(\beta \supset \alpha) \qquad \text{df}(\equiv)_1$$

$$A \equiv B \stackrel{\text{def}}{=} (A \supset B)\,\&(B \supset A) \qquad \text{df}(\equiv)_2$$

Assuming that u and v are distinct and that u has no free occurrence in α or B and v has no free occurrence in A, the system has:

$$(\exists u)Au \supset \alpha \stackrel{\text{def}}{=} (u)(Au \supset \alpha) \qquad \text{dfs(P)}_1$$

$$\alpha \supset (u)Au \stackrel{\text{def}}{=} (u)(\alpha \supset Au) \qquad \text{dfs(P)}_2$$

$$(u)Au \supset \alpha \stackrel{\text{def}}{=} (\exists u)(Au \supset \alpha) \qquad \text{dfs(P)}_3$$

$$\alpha \supset (\exists u)Au \stackrel{\text{def}}{=} (\exists u)(\alpha \supset Au) \qquad \text{dfs(P)}_4$$

$$(\exists u)Au \supset (\exists v)Bu \stackrel{\text{def}}{=} (u)(\exists v)(Au \supset Bv) \qquad \text{dfs(X)}_1$$

$$(u)Au \supset (v)Bv \stackrel{\text{def}}{=} (v)(\exists u)(Au \supset Bv) \qquad \text{dfs(X)}_2$$

$$(\exists u)Au \supset (v)Av \stackrel{\text{def}}{=} (u)(v)(Au \supset Bv) \qquad \text{dfs(X)}_3$$

$$(u)Au \supset (\exists v)Bv \stackrel{\text{def}}{=} (\exists u)(\exists v)(Au \supset Bv) \qquad (\text{dfs(X)}_4)$$

This system is complete with respect to quantification theory, see [12].

There was a major hurdle, however. Without general propositions, Russell cannot generate the existence theorems in substitution that are needed to generate arithmetic. Axiom schema (S17) is now such that its formulas A must be quantifier-free. Consider the following comprehension theorem schemas derived from (S17):

$$(\exists p,\ a)(x)(p/a; x.\equiv .\{A\}), \qquad (\text{CP})^1$$

where p and a are not free in A.

$$(\exists p,\ a,\ b)(x,\ y)(p/a, b; x, y.\equiv .\{A\}), \qquad (\text{CP})^2$$

where p, a and b are not free in A. These will be well-formed only if the formula A in them is quantifier-free. In the system of "On 'Insolubilia'" there is no term $\{A\}$ unless the formula A is quantifier-free. This is a severe limitation. For example, it will no longer be possible to generate the existence theorems needed for the theorem $\Lambda^{(o)} \in 0^{((o))}$. One would need,

$$(\exists s,\ t,\ w)(s/t, w \approx_{r,c} \{0\,(r,\ c)\})$$

and this is no longer follows from $(CP)^2$. Recall that,

$$0\,(r,\ c) \stackrel{\text{def}}{=} (z) \sim (z \in \iota(p/a)[p/a;\ x \equiv_x r/c;\ x]).$$

In the substitutional theory of "On 'Insolubilia'" there is no term $\{0(r, c)\}$ because it is not possible in the revised theory to nominalize a general formula.

To mitigate the affect of the abandonment of general propositions, Russell offered what amounts to auxiliary axioms for comprehension. For example, he has:

$$(\exists p,\ a)(x)(\exists q)(p/a;\ x!q.\ \&\ .q \equiv A), \qquad 1906(\text{Aux})^1$$

where A is any wff (quantifier-free or otherwise) in which p, a are not free.

$$(\exists s,\ t,\ w)(r,\ c)(\exists q)(/t,w;r,c!q.\ \&\ .q \equiv A), \qquad 1906(\text{Aux})^2$$

where A is any wff (quantifier-free or otherwise) in which s, t, w are not free. (It is in virtue of Russell's contextual definitions of quantified-formulas flanking logical particles that the formula A in these comprehension axiom schemata can contain quantifiers.) By means of these new auxiliary comprehension principles, arithmetic can be recovered in the 1906 substitutional theory. It must be understood that Russell's "mitigating" comprehension principles are not "reducibility" principles of a system espousing a hierarchy of orders of propositions. In "On 'Insolubilia'" there are no orders of propositions at all.[15] So the revised theory preserves Russell's doctrine that any proper calculus for the science of logic must have only one style of variables—viz., individual variables. Unfortunately, however, Russell overlooked the fact that the "mitigating" comprehension principles will resurrect the po/ao paradox! The following is an instance of 1906 $(\text{Aux})^1$:

$$(\exists t,\ w)(x)(\exists s)(t/w;\ x!s.\ \&\ .(s.\ :\equiv:\ .\ x = \{p \supset a\}.\ \&\ .p/a;\ x!r\ :\supset_{p,a,r}:\ \sim r)).$$

The paradox goes through. Russell had to admit that his new mitigating axioms were too strong. Without them, however, the system of "On 'Insolubilia'" with its abandonment of general propositions fails to be strong enough to recover arithmetic.

7. A New Approach to 'Insolubilia'

The abandonment of general propositions in "On 'Insolubilia'" had not yielded a solution of the po/ao paradox which could recover arithmetic. In 1907 Russell entertained the possibility of ramifying the substitutional theory by introducing order indexed variables in an effort to avoid the po/ao paradox (and its variants). Russell began to investigate what formal system would result if the substitutional theory were fitted with order indexed variables. General propositions would be assumed again,

[15] Cocchiarella in [2] makes this point. Hylton attributes orders of propositions to the system Russell set out in "On 'Insolubilia'." See [7].

and propositions would be split into orders on the basis of the kind of generality they "involve" or "presuppose". Aspects of this idea surfaced in Russell's paper "Mathematical Logic as Based on the Theory of Types." With substitution retrofitted with order indexed variables, axioms of propositional reducibility can be added without fear of reviving the po/ao paradox.

We saw that the pristine substitutional theory begun in 1905, proxies a simple type theory of attributes. When the substitutional theory is fitted with order indices, it will reconstruct a ramified type-theory of attributes. Russell acknowledges that the language of predicate variables fitted with order/type indices is convenient because it supports the introduction of ordinary class symbols. But the underlying foundation of ramified types in "Mathematical Logic" is the substitutional theory. "Mathematical Logic" took a long time at the press, and when it finally appeared in 1908, Russell thoughts had changed. In a letter to Hawtrey of 1907, we find Russell explaining that the po/ao paradox had "pilled" the substitutional theory and that he was never satisfied with the patches he devised.[16] By 1908, the era of substitution was over. Its successor was the ramified type-theory of *Principia* with it reliance on reducibility and infinity axioms.

I believe that the approach in "On 'Insolubilia'" was on the right track. Russell needs to drop his mitigating axioms and find new auxiliary axiom schemata.[17] The way to find them is to observe that propositions, unlike the formulas of the language of substitution, can have infinitely many constituents. Some propositions contain finitely many constituents. As we shall see, it is provable that there are infinitely many of them. Indeed, it seems natural to imagine that there are exactly $\aleph_0$ many such propositions. Since propositions are finely individuated there will be $2^{\aleph_0}$ many propositions containing exactly $\aleph_0$ constituents (each of which has finitely many constituents).[18] The number of propositions that have exactly $2^{\aleph_0}$ constituents is 2 to the $2^{\aleph_0}$ and so on. Let us introduce into the language of substitution primitive predicate expressions of the form $C^n(x)$. In the intended interpretation, $C^0(x)$ says that x has finitely many consituents, $C^1(x)$ says that the entity x contains exactly $\aleph_0$ many constituents, $C^2(x)$ says the entity x contains exactly $2^{\aleph_0}$ constituents, and so on. Define as follows:

$$(x_n)Ax_n \stackrel{\text{def}}{=} (x)(C^n(x) \supset Ax)$$
$$(\exists x_n)Ax_n \stackrel{\text{def}}{=} (\exists x)(C^n(x)\ \&\ Ax).$$

We'll need to add the following axiom schema:

$$C^{n_1}(\beta_1), \&, \ldots, \& .C^{n_j}(\beta_j) :\supset: C^m(\{A\}),$$

[16] Letter to Hawtrey, dated 22 January 1907.

[17] I made some preliminary suggestions in [11].

[18] It may be, however, that since propositional structure is well-ordered the number is $\aleph_1$. My thanks to David McCarty for a helpful discussion of this matter.

where $\beta_1, \ldots, \beta_j$ are all the terms free in A and $m = max(n_1, \ldots, n_j)$. The needed mitigating axioms then include the following:

$$(\exists p_1,\ a_1)(x_0)(\exists q)(p_1/a_1; x_0!q.\,\&\,.q \equiv A), \qquad 2001(\text{Aux})^1$$

where A is any wff (quantifier-free or otherwise) in which p, a are not free.

$$(\exists s_2,\ t_2,\ w_2)(r_1,\ c_1)(\exists q)(_2/t_2, w_2; r_1, c_1!q.\,\&\,.q \equiv A), \qquad 2001(\text{Aux})^2$$

where A is any wff (quantifier-free or otherwise) in which s, t, w are not free.

It is easy to see that these axiom schemata should be viewed as truths of the logic of propositions. Suppose there are only $\aleph_0$ many individuals that have finitely many constituents. No matter what formula $A(\xi)$ is, there is a proposition $\boldsymbol{p}$ containing exactly $\aleph_0$ many constituents, viz.,

$$\{\boldsymbol{a} = x.\vee.\boldsymbol{a} = x'.\vee.\boldsymbol{a} = x''.\vee.\boldsymbol{a} = x'''.\vee.\boldsymbol{a} = x''''\ldots\},$$

where the entity $\boldsymbol{a}$ contains exactly $\aleph_0$ many constituents, and where x', x'', x''', and so on, are all and only the entities with finitely many constituents that satisfy the formula.[19] But then we have $C^1(\boldsymbol{p})$ and $C^1(\boldsymbol{a})$ and $(x_o)(\exists q)(\boldsymbol{p}/\boldsymbol{a}; x_0!q.\,\&\,.q \equiv A(x_o))$. Hence we have:

$$(\exists p_1, a_1)(x_0)(\exists q)(p_1/a_1; x_0!q.\,\&\,.q \equiv A)$$

and this is just $2001(\text{Aux})^1$. The same point holds as we ascend types. Instances of the axiom schemata $2001(\text{Aux})^1$, $2001(\text{Aux})^2$, etc., will all count as truths of the pure logic of propositions. In this way, substitution can avoid the axiom (schemata) of reducibility that undermines logicism in *Principia Mathematica*.

It might be objected that the C^n predicates which are adopted as part of the system's primitive constants take the notion of cardinal number as primitive. But observe that the C^n predicates do not *themselves* generate the notion of the cardinal number of a class. That notion is given its usual Frege/Russell construction. It is our semantic interpretation of the C^n predicates that takes them refer to the cardinality of a proposition. There is no axiom of the system that assures that some propositions have greater than finite cardinality. That result is a theorem of the system, derived from the po/ao paradox itself. That is,

$$(x)(C^n x \supset \sim C^m x),$$

where $m \neq n$ is a theorem schema of the system. The proof is as follows. Suppose,

$$\iota(p/a)[p/a; x \equiv_x C^1 x]\,\text{SIM}\,\iota(p/a)[p/a; x \equiv_x C^0 x].$$

That is, suppose there were a function from the class of all entities x that are $C^1 x$ onto the class of all entities x that are $C^0 x$. The classes are then similar. In substitutional

[19] The language of the substitutional theory, of course, is finitary. But there are curious ontological questions about infinite propositions composed of infinite iterations of a binary relation. Many thanks to David Kaplan for helpful discussions at the conference about whether an infinitary relation would be needed as the universal which occurs predicatively in such a proposition.

notation this is:

$$(\exists h,\ d,\ e)(\text{1-1funct}(\iota(s/t, w)[s/t, w \approx^{(0,0)} h/d, e]) \&$$
$$\iota(p/a)[p/a; x \equiv_x \text{Dom}^{(0,0)}(hde)(x)] = \iota(p/a)[p/a; x \equiv_x C^1 x] \&$$
$$\iota(p/a)[p/a; x \equiv_x \text{Rng}^{(0,0)}(hde)(x)] = \iota(p/a)[p/a; x \equiv_x C^0 x]).$$

(The substitutional definitions of domain, range, and functionality are as expected.) After an existential instantiation, let us abbreviate as follows:

$$Gx_0 \stackrel{\text{def}}{=} (\exists m,\ n)(\langle p, m\rangle \in \iota(s/t, w)[s/t, w \approx^{(0,0)} h/d, e] \&$$
$$\langle a, n\rangle \in \iota(s/t, w)[s/t, w \approx^{(0,0)} h/d, e] \& x_0 = \{m \supset n\}).$$

Then we can formulate a version of the po/ao paradox as follows:

$$(\exists t_1,\ w_1)(x_0)(\exists q)(t_1/w_1; x_0!q. \& .(q. :\equiv: .\ Gx_0. \& .p/a; x_0!r ::\supset_{p,a,r}: \sim r)).$$

Contradiction results. Accordingly, there can be no function from the class of all entities x that are $C^1 x$ to the class of all entities x that are $C^o x$. We can go on to prove the following:

$$\iota(p/a)[p/a; x \equiv_x C^o x] < \iota(p/a)[p/a; x \equiv_x C^1 x].$$
$$(x)(C^1 x \supset \sim C^0 x).$$

By similar considerations, we know

$$\iota(p/a)[p/a; x \equiv_x C^n x] < \iota(p/a)[p/a; x \equiv_x C^{n+m} x].$$
$$(x)(C^n x \supset \sim C^{n+m} x).$$

For this reason it is natural for the C^n predicates to be semantically interpreted as cardinality predicates. But the differences in cardinality are theorems, not axioms of the system.

Our reformulation of Russell's substitutional theory avoids the difficulties that undermine *Principia's* logicism. The system avoids *Principia's* infinity axiom. In the system, we can capture a universal class of all *entities* whatsoever,

$$\iota(p/a)[p/a; x \equiv_x \text{V}x].$$

(I put: $\text{V}\xi \stackrel{\text{def}}{=} \xi = \xi$.). It is easy to prove that this class is Dedekind infinite (i.e., it is similar to one of its proper subsets). There is also the class of all entities with finitely many constituents:

$$\iota(p_1/a_1)[p_1/a_1; x_0 \equiv_{x_0} \text{V}x_0].$$

This is also Dedekind infinite. We can prove that any class that is Dedekind infinite is infinite in Frege's sense of not being a member of any natural number. Thus we know that the class $\iota(p_1/a_1)[p_1/a_1; x_0 \equiv_{x_0} \text{V}x_0]$ is infinite in Frege's sense. To be sure, this leaves it open whether there are exactly $\aleph_0$ many propositions with finitely

many constituents. There could be more! This won't jeopardize our argument for the auxiliary axiom schemata of the revised system of substitution. The auxiliary axiom schemata will be valid in the domain of propositions no matter what infinite cardinality the set of entities with finitely many constituents is. Thus we have avoided both the infinity axiom and reducibility axiom (schemata) of *Principia*. These were the two main hurdles. It would seem then, that in the modified substitutional system just outlined, the 'Insolubilia' plaguing logicism—Russell's paradoxes of classes, attributes, and his po/ao paradox—are finally solved. Russell gave up on substitution too quickly. The "hoary difficulties of the one and the many," as Russell called them, are indeed solved by the substitutional theory, just as Russell had hoped.

References

[1] Church, Alonzo: 1984. Russell's theory of the identity of propositions. *Philosophia Naturalis* 21: 513–22.

[2] Cocchiarella, Nino: 1980. The development of the theory of logical types and the notion of a logical subject in Russell's early philosophy. *Synthese* 45: 71–115.

[3] de Rouilhan, Philippe: 1996. *Russell et le cercle des paradoxes*. Paris: Presses Universitaires de France.

[4] Grattan-Guinness, Ivor: 1974. The Russell archives: Some new light on Russell's logicism. *Annals of Science* 31: 387–406.

[5] Grattan-Guinness, Ivor: 1977. Bertrand Russell's manuscripts: An apprehensive brief. *History and Philosophy of Logic* 6: 53–74.

[6] Grattan-Guinness, Ivor: 1977. *Dear Russell—Dear Jourdain*. London: Duckworth.

[7] Hylton, Peter: 1980. Russell's substitutional theory. *Synthese* 45: 1–31.

[8] Lackey, Douglas: 1973. *Essays in Analysis by Bertrand Russell*. New York: George Braziller.

[9] Landini, Gregory: 1989. New evidence concerning Russell's substitutional theory of classes. *Russell* 9: 26–42.

[10] Landini, Gregory: 1998. *Russell's Hidden Substitutional Theory*. New York: Oxford University Press.

[11] Landini, Gregory: 1998. Russell's intensional logic of propositions: A resurrection of logicism? In: F. Orilia and W. J. Rappaport (eds.), *Thought, Language, and Ontology: Essays in Honor of Hector Neri Castañeda*, Amsterdam: Kluwer, 61–93.

[12] Landini, Gregory: 2000. Quantification theory in *9 of *Principia Mathematica*. *History and Philosophy of Logic* 21: 57–78.

[13] Peano, Guisseppe: 1906. Addition. *Revista de Mathematica* 8: 143–57.

[14] Pelham, Judy and Urquhart Alasdair: 1994. Russellian propositions. In: D. Prawitz *et al.* (eds.), *Logic, Methodology and Philosophy of Science IX*, Proceedings of the Uppsala conference, 307–26.

[15] Potter, Michael: 2000. *Reason's Nearest Kin*. Oxford: Oxford University Press.

[16] Quine, Willard V.: 1967. Introduction to Russell's *Mathematical Logic as Based on the Theory of Types*. In: J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic 1879–1931*, Harvard: University Press, 150–152.

[17] Russell, Bertrand: 1903. *The Principles of Mathematics*. Cambridge. References are to the second edition London: W. W. Norton, 1967.

[18] Russell, Bertrand: 1905. On denoting. *Mind* 14: 479–93.

[19] Russell, Bertrand: 1905. On substitution. Unpublished manuscript dated 22 December 1905. Russell Archives, McMaster University, Hamilton, Ontario, Canada. In [30].

[20] Russell, Bertrand: 1905. On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society* 4: 29–53. The paper was read to the Society on 14 December 1905. Pagination is to the paper as reprinted in [29]: 135–164.

[21] Russell, Bertrand: 1905. On the substitutional theory of classes and relations. Received by the London Mathematical Society on 24 April 1906, and read before the Society on 10 May 1906. In [29]: 165–89.

[22] Russell, Bertrand: 1906. On substitution. Unpublished manuscript of April–May 1906 (Russell Archives). In [30].

[23] Russell, Bertrand: 1906. Les paradoxes de la logique. *Review de Métaphysique et de Morale* 14: 627–59. Pagination is to the English manuscript *'Insolubilia' and Their Solution by Symbolic Logic* in [29]: 190–214.

[24] Russell, Bertrand: 1906. The paradox of the liar. Unpublished manuscript of September 1906 (Russell Archives, McMaster University, Hamilton, Ontario, Canada). In [30].

[25] Russell, Bertrand: 1908. Mathematical logic as based on the theory of types. *American Journal of Mathematics* 30: 222–62.

[26] Russell, Bertrand: 1924. Logical atomism. In: R. Marsh (ed.), *Logic and Knowledge Essays 1901–1950*, London: Allen & Unwin, 1977. 321–343.

[27] Russell, Bertrand. 1927. *Philosophy*. New York: W. W. Norton & Company, Inc.

[28] Russell, Bertrand: 1959. *My Philosophical Development*. New York: Simon and Schuster.

[29] Russell, Bertrand: 1973. *Essays in Analysis by Bertrand Russell*. Edited by D. Lackey. New York: Braziller.

[30] Russell, Bertrand: forthcoming. *The Collected Papers of Bertrand Russell, vol. 5*. London: Routledge.

[31] Russell, Bertrand and Couturat, Louis: 2001. *Bertrand Russell correspondence sur la philosophie, la logique et la politique avec Louis Couturat (1997–1913)*. Edited by Anne-Françoise Schmidt. Paris: Edition Kimé.

[32] Whitehead, Alfred North and Bertrand Russell: 1910. *Principia Mathematica*. Cambridge, 1910.

Department of Philosophy
University of Iowa
Iowa City, IA 52242-1408
USA

E-mail: gregory-landini@uiowa.edu

Substitution and Types: Russell's Intermediate Theory

Philippe de Rouilhan

Abstract. Russell went through three "substitutional" theories on his way to the so-called ramified theory of types. The first was so strong as to be inconsistent; the second was consistent, but too weak. The third is mentioned in passing in "Mathematical Logic as Based on the Theory of Types", and abandoned in favor of the ramified theory of types for the sake of "practical convenience". However, it is doubtless the most fascinating theory Russell ever came up with.

1. Introduction

We now know what went wrong in the article "On the Substitutional Theory of Classes and Relations" that Russell read on May 10, 1906 before the London Mathematical Society and why he decided not to publish it some time later. The theory presented in it was contradictory ([2], [1], [5]: chap. V, [3]: chap. VIII), and Russell had finally realized that ([3], *ibid.*). The elimination of functions and of classes and relations[1] as entities and their contextual redefinition as logical fictions had indeed sheltered the theory from the paradoxes to which Russell had long known these entities to give rise, but the propositions were now producing a new paradox. Not a semantical or epistemological paradox, but a purely logical paradox. As early as the summer of 1902, Russell had himself discovered such a paradox involving the notion of proposition that afflicted the "doctrine of types" that he had just developed to escape from the other logical paradoxes. Curiously, however, he did not immediately realize some four years later that the substitutional theory was to meet exactly the same fate ([7]: appendix B, in particular §500; [8]: *in fine*).

Once he knew what was going on, Russell had to solve the paradox of propositions. The first idea that could have come to mind was to eliminate propositions

[1] In those days, Russell applied the classical *intension/extension* distinction to "classes" as well as to "relations". It is, therefore, both classes and relations in extension or in intension that are involved here. But from 1908 on, it would no longer be a question of classes and relations *in intension* since Russell very likely considered them to be doing double duty with the functions themselves. To avoid long, awkward sentences, I shall speak of "classes and relations" without being any more explicit, since taking the dates into consideration suffices to remove the ambiguity.

altogether. But the vicious circle principle which Poincaré had just set forth didn't require such a drastic move. It only required that one give up postulating propositions containing apparent variables[2], viz. general propositions. Russell therefore took pains to reconstruct logic in the form of a new substitutional theory: one without classes, relations, functions, and now, without general propositions. This logic had to be weak enough to be consistent, but strong enough for one to be able to reconstruct classical mathematics within it. And he did not manage to construct it.[3]

This second substitutional theory had to be a logic that was both predicative (in keeping with the vicious circle principle) and universal (only logically recognizing a single kind of being and therefore only mobilizing a single sort of variable). Through the iterative character of their logical reconstruction, the functions, and likewise the classes and relations, of the first substitutional theory formed a hierarchy. According to the second theory, it was to be the same for general propositions, which had to form a certain hierarchy of logical fictions constructed on the basis of a real world exclusively composed of individuals (particulars or universals) and of elementary propositions. So Russell failed to achieve his goals.

He would soon give up. He would give up, at least temporarily, on universality. Unable to *construct* the hierarchy of general propositions, he would *give* it to himself. More specifically, he would grant himself a certain hierarchy of individuals and propositions, elementary or general, satisfying the vicious circle principle. The primitive logical notion of substitution would be retained, enabling one, in principle, to construct the fictional hierarchy of functions, then that of classes and relations, on the basis of that real hierarchy. Russell would mention this third theory, both substitutional and typed, in the 1908 article, "Mathematical Logic as Based on the Theory of Types", but would abandon it in turn, finding it "technically inconvenient"([11], [18]: 77), in favor of his famous "theory of types" known today as the "ramified theory of types". The primitive notion of substitution would disappear, and the hierarchy of functions would be henceforth treated as a given; the only thing left would be to construct that of classes and relations.

So in between the second substitutional theory and the ramified theory of types came the "intermediate theory", as I call it. Russell himself passed over so quickly that it has scarcely held the attention of scholars. Yet, it is doubtless the most fascinating theory Russell ever came up with.[4]

[2]The notion of variable is to be understood here in a Russellian, extra-linguistic, sense.

[3][5] and [3] (the latter taking into account unpublished manuscripts of the time) give divergent explanations of this point. This is not the place to enter into a debate about this.

[4]The following sections overlap quite a bit with my [5]: chap. VI, section A.

2. Predicativity without Universality: Types of Entities

The vicious circle principle and its technical version, which appeared in 1906 [9], reappeared without significant change in 1908. Admittedly one can detect a few variations in the heuristic versions,[5] but the technical version is practically the same ('involves' was replaced by 'contains'):

> (VCP1) Whatever contains an apparent variable must not be a possible value of that variable. ([11]: 237, [18]: 75)

Significant change definitely took place between 1906 and 1908. It lies in the abandonment of universalism, temporary, to be sure, but abandonment all the same. It is fitting to take into account this abandonment when evaluating the consequences of VCP for the construction of logic. Like Russell, let us call "type" the range of significance of some propositional function of one argument, in other words, the domain of possible values of a variable.[6] In view of the requirement of universality, the VCP had done away with the objects containing apparent variables; now that this requirement was lifted, it merely relegated them to another type. Henceforth there could well be entities containing apparent variables, but they would simply be of a different type than the entities that these variables could take as values, which Russell claims as an immediate consequence of VCP1:

> (VCP2) [W]hatever contains an apparent variable must be of a different type from the possible values of that variable. ([11]: 237, [18]: 75)

The universe, therefore (if one may still speak of *uni*verse), would in some way be divided into types, and the variables themselves would be divided up according to the types of their values. But this division would not be just any division, it would respect a certain number of constraints. I shall indicate two of them. The first hangs on the very idea of the vicious circle principle. This principle says that an entity b which contains an apparent variable must be of a different type from any possible value a of that variable, but the idea is that the same must hold for any entity c containing an apparent variable able to take b as a value, i.e., c must be of a different type from a,

[5] "[A]ll our contradictions have in common the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself. This leads us to the rule: 'Whatever involves *all* of a collection must not be one of the collection'; or, conversely: 'If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total' " ([11]: 225, [18]: 63); and a footnote then states in reference to this last word: "When I say that a collection has no total, I mean that statements about *all* its members are nonsense". A few pages later: "These fallacies, as we saw, are to be avoided by what may be called the 'vicious circle principle'; i.e., 'no totality can contain members defined in terms of itself'. This principle, in our technical language, becomes: 'Whatever contains an apparent variable must not be a possible value of that variable' " ([11]: 237, [18]: 75).

[6] "A *type* is defined as the range of significance of a propositional function, i.e., as the collection of arguments for which the said functions has values. Whenever an apparent variable occurs in a proposition, the range of values of the apparent variable is a type, the type being fixed by the function of which 'all values' are concerned" ([11]: 237, [18]: 75).

and the same must also go for any entity d that contains an apparent variable able to take c as a value, i.e., d must be of a different type than a, and so on *ad infinitum*. The differences of types would be hierarchical differences. When he stated VCP2, Russell immediately added: "we will say that it is of a *higher* type". This is not merely a matter of terminology. The second constraint is different in origin, perhaps a concern for simplicity (Russell did not give any explanation), and the VCP2 formulation of the vicious circle principle is connected with it. The vicious circle principle did not in itself require that the types be pairwise disjoint. In fact, they would be. They would form a partition (in the mathematical sense of the term) of the universe. Any entity would belong to one, and only one, type. I shall refer to this constraint as the "principle of disjunction (of types)".[7]

3. On the Intended Interpretation of the Intermediate Theory

In the intermediate theory an ontology of individuals and propositions (elementary or general) arranged in accordance with a hierarchy of types, as stipulated by the vicious circle principle and the principle of disjunction, is taken for granted and one has to construct functions on this basis, then classes and relations.

Several hierarchies of individuals and propositions can *a priori* satisfy the vicious circle principle and the principle of disjunction. One expects to see Russell adopt the one in accordance with which he thinks the general propositions in question here would have been arranged if he had succeeded in reconstructing them within the second substitutional theory. At the time, in 1906, Russell only expressly envisioned as general propositions those for which the apparent variables were variables of entities. But, one must see that if he had succeeded in logically reconstructing these propositions, he would have been led to consider new general propositions in which variables taking the preceding ones as values would have figured. And this process would have had to go on indefinitely. It is the hierarchy in accordance with which all these propositions would have been arranged of their own accord that is at stake here.

However, this hierarchy, at least as Russell could have imagined constructing it, and the one that he provided himself with in 1908 differ in at least two ways. In 1906, there was an element of typological complexity that fell out in 1908. In 1906, Russell thought that the general propositions that he expressly took into account would be of different types depending on the number of their apparent variables, see ([9]: 643, 647) and ([19]: 207, 211). In 1908, it seems he fortunately no longer thought that, but he didn't comment on the change.[8] Given this simplification, one would expect

[7] [12]: 95, formulated a principle that amounts to the same thing and which corresponds to Tarski's "first principle of the theory of semantical categories" ([20], [22]: 216).

[8] Curiously, this thought comes up again in *Principia* as a *theoretical* requirement with which it is *practically* useless to comply ([12]: 133, 162).

Russell to reason in the following fashion: The entities that had been countenanced in the second substitutional theory, namely individuals and elementary propositions, would be of the first type, let us say of type 1; the propositions containing variables of type 1, but of no higher type, would be of type 2; the propositions containing variables of type 2, but of no higher type, would be of type 3; etc.

However, that is not exactly what Russell did. The hierarchy that he assumed differed from the preceding one (with all that that implies) on one point: the type of elementary propositions. Above, the elementary propositions would be of the same type as the individuals; they would be of type 1. For Russell, only individuals were of type 1. As for elementary propositions, they were of the same type as the general propositions containing variables of type 1, but of no higher type; they were of type 2. This time it was an element of ontological complexity affecting the structure of the elementary propositions that fell out. The description of higher types could *literally* remain the same. Let us add that Russell drew in a notion that would acquire full significance only within the ramified theory of types, but which here did double duty with the notion of type, viz. the notion of *order*. In the Russellian hierarchy, propositions are attributed a type *and* an order: propositions of type 2 are of order 1, those of type 3 of order 2, etc.[9] It is natural to extend the notion of order to individuals by attributing them order 0.

Russell did not advance an argument of ontological simplicity in defense of his decision concerning the type of elementary propositions, but did give an argument to the effect that if one wants to avoid a vicious circle, no proposition must be of the same type as an individual.[10] However, he went too far; as he had in fact already recognized in 1906, it was enough that no *general* proposition was. But being a matter of elementary propositions, in any case, something happened in 1907, even before Russell became involved in the theory of types, which deserves to be mentioned. *Grosso modo*, in that year the possibility occurred to Russell of accommodating certain arguments militating against the notion of proposition in general, and of reducing elementary propositions themselves to logical fictions. It is therefore not surprising that in the intermediate theory they were no longer of the same type as individuals. But let us be more specific, because that is not, as I see it, the whole story.

[9] "*[I]ndividuals* (…) form the first or lowest type. (…) Elementary propositions together with ones containing only [variables of] individuals as apparent variables will be called *first-order propositions*. These form the second logical type. (…) We can (…) form new propositions in which [variables of] first-order propositions occur as apparent variables. These we will call *second-order propositions*; these form the third logical type. (…) The $n+1$th logical type will consist of propositions of order n, which will be such as contain [variables of] propositions of order $n-1$, but of no higher order, as apparent variables. The types so obtained are mutually exclusive" ([11]: 237–238, [18]: 76–77 additions in square brackets are mine).

[10] "By applying the process of generalization to individuals occurring in elementary propositions, we obtain new propositions. The legitimacy of this process requires only that no individuals should be propositions. (…) Hence in applying the process of generalization to individuals we run no risk of incurring reflexive fallacies". ([11]: 237–238, [18]: 76)

4. Looking back: The Theory of Judgment of 1907 and the Elimination of Elementary Propositions

Up until 1907, Russell had countenanced propositions in his ontology albeit reluctantly. They provided a way of: 1) accounting for the meaningfulness of sentences independently of their truth value; 2) assigning a bearer for these truth values; and 3) assigning an object[11] to propositional attitudes[12], to belief, for example, independently of their rightness or wrongness. But there was something shocking about granting false propositions the same ontological status as true propositions, as facts. In his 1907 article, "On the Nature of Truth" [10], Russell reexamined the notion of proposition from the point of view of the theory of belief. Elementary propositions were called into question just as much as general propositions.

Two theories were envisioned. According to the first, when belief is correct, or true, it is so because there is a fact corresponding to it as its object; when it is mistaken, or false, it is because there is nothing of the kind. According to the second theory, belief always has an object, and this object is a *proposition*: a true proposition, or *fact*, if the belief is correct; a false proposition, or *fiction* (in the ordinary sense, not in the logical sense), if the belief is mistaken. The first theory has simplicity and common sense on its side, and it displays a sort of ontological parsimony which speaks in its favor. But going against it is the fact that it provides no answer to the following question: "What do we believe when our belief is mistaken?" ([10]: 45). If our belief is about nothing, it is in no way a belief. That is the objection that leads from the first theory to the second.

Russell sought to amend the first theory in order to escape the second. He did it by exposing one of its presuppositions. This presupposition is that belief is a "*single* state of mind", meaning a "*single* complex idea", and therefore a "*single* thing". If this were the case, then this mental state, this complex idea, should always have a *single* object, a certain entity (specifically a proposition), and one would be led, as the second theory in fact does, to grant the same ontological status to "non-facts" (the expression is Russell's) as to facts, to objective falsehoods as to objective truths, to false propositions as to true propositions.

Russell denied that this was the case. He maintained that the belief that *aRb*, for example, is not *one* (complex) idea, but (a complex of) *several* ideas: the idea of a, the idea of b, the idea of R. Each of these ideas has its own object: a, b, R; but there is no need to postulate, or hypostasize, in all cases something like the object *aRb* of the corresponding belief. Whether this object exists (whether *aRb* is a fact—case of correct belief) or not (case of mistaken belief); whether the belief, in this sense, is "objective" or not, does not matter: the principle "no idea without object" still holds, and belief without object (mistaken belief) remains in a full-fledged belief.

Apart from the common sense argument that has it that in the world there are no "non-facts", objective falsehoods, or false propositions, Russell set forth an argument

[11] In this section, "object" will be taken in the narrow sense, according to which an object is an entity.

[12] As Russell would say in 1918, cf. [18]: 227.

in defense of his theory of belief that is worth discussing. In his eyes, in 1906, not only the propositional paradox, but also epistemological versions of the Liar's paradox, had shown that not all propositions were entities. Similarly a paradox analogous to the latter now showed that not all beliefs (mental states, i.e., complexes of ideas in which they consist) were entities. It was a matter of a man who, only having mistaken beliefs otherwise, believes that all his beliefs are mistaken, including this one. If this belief is also mistaken, then indeed all his beliefs are mistaken, and he is right to believe that: so the belief in question is correct after all. Conversely, if that belief is correct, then not all his beliefs are mistaken, and so he is wrong to believe the contrary: hence the belief in question is mistaken. In 1906, only general propositions had been dismissed, in keeping precisely with what the vicious circle principle required. In the same way, now, only general beliefs, as it were, *qua* entities had to be so. But the new theory of belief was primarily designed to satisfy the requirement of ontological parsimony: all false propositions had to disappear, therefore also all mistaken beliefs qua entities, and hence, by parity, all beliefs qua entities.

In 1910, in the last chapter of the *Philosophical Essays* entitled "On the Nature of Truth and Falsehood" [13], Russell would return to the theory of belief outlined in 1907 and make it more rigorous. He would no longer be interested in the ontological status of beliefs, but rather in the existence of their objects. And he would not merely state that there are no more *false* propositions, but no more *propositions* at all. For if propositions as such are defined as objects of belief, if mistaken belief is void of object, and if correct belief must be analyzed in itself like mistaken belief, then beliefs in themselves are void of object and there are not, properly speaking, any *propositions*. That does not mean that there are no correct beliefs and thus no facts; the latter are simply no longer the objects of the former: they are not *propositions*. The theory of belief, then, that Russell ultimately espoused can be summed up as follows: All belief is void of object, and there are no propositions; propositions are only logical fictions; a belief is not a relation of a subject to *one* object, but to *several*; the latter constitute a unit, a *fact*, if the belief is correct, and they do not constitute anything at all if it is mistaken. (Naturally, what Russell says about belief is presumed to apply to all propositional attitudes, and go with a theory of the sentence meaning and of bearers of truth values that also dispenses with propositions.)

Let us establish, as Russell did not do, the connection with the plan for a logic in the state in which he had left it in 1906. The ontology had been reduced to elementary propositions and their ultimate constituents, individuals. In 1907, elementary propositions themselves disappeared to leave room for elementary facts.[13] From the 1906 perspective, the job of logical reconstruction had already seemed impossible. *A fortiori*, it appeared so in 1907. There is definitely nothing surprising about the 1908 retreat back to a predicative, but non-universal logic. What might surprise us is that elementary facts and their ultimate constituents, individuals, are not exactly the entities of the lowest type in the intermediate theory. It is clear that they are not, since

[13]The new elementary facts must not be identified with the old ones, with the ex-propositions true elementary.

only "individuals" are found in the lowest type, and "[w]e may define an individual as something destitute of complexity" ([11]: 238, [18]: 76). That elementary propositions disappeared in 1907 is in accordance with their being found in the second type of the intermediate theory; their replacement in 1907 by elementary facts could have led Russell to put the latter in the first type of that theory. The argument of ontological simplicity may have operated. Perhaps Russell shrank from the idea of having an entity (an elementary fact) figure in the first type that in some way would have been found again in the second (as a true elementary proposition). Be that as it may, that would not have left any more of a trace in the intermediate theory than it would do in the ramified theory of types if he had developed the first as he would do the second.

5. Towards the Formalization of the Intermediate Theory: The Language L(IT)

Below, I propose a formal language, L(**IT**), to which one is naturally led from the language of the first substitutional theory when, like Russell: 1) one seeks to satisfy the vicious circle principle, no longer by giving up general propositions, but by arranging them in accordance with a certain typology; 2) one chooses a typology satisfying the principle of disjunction; 3) instead of putting elementary propositions in the same type as individuals, one includes them in the same type as first-order propositions. It is to this very language L(**IT**) that one is led if to these considerations one adds those of the preceding section. In my exposition of this language, I shall often employ the notion of type when it is a question of indeterminate entities, individuals or propositions, and the notion of order when it is more precisely a question of propositions.

First, some introductory remarks about L(**IT**) and its interpretation. L(**IT**) will be a language with infinitely many sorts of variables: variables of individuals, 'x_1^0', 'x_2^0', 'x_3^0', …; variables of first-order propositions, 'x_1^1', 'x_2^1', 'x_3^1', …; variables of second-order propositions, 'x_1^2', 'x_2^2', 'x_3^2', …; etc. Alongside the variables there may be an array of constants of individuals, 'a_1^0', 'a_2^0', 'a_3^0', …; constants of first-order propositions, 'a_1^1', 'a_2^1', 'a_3^1', …; constants of second-order propositions, 'a_1^2', 'a_2^2', 'a_3^2', …; etc.

The well-formed symbols, or well-formed expressions, will be defined recursively at the same time as an order is assigned to them and the notion of real (apparent respectively) occurrence of a variable in such a symbol is defined. For any values (of appropriate type) assigned to the variables for the real occurrences that they possess there, these symbols will correspond to entities (of a certain type depending on the symbol considered in each case) and for this reason will be called "entity symbols", or "complete symbols". If these entities are propositions, they will be called "proposition symbols", or "formulas". The formulas without real occurrence of variables, corresponding to definite propositions, will be "sentences".

Among the primitive symbols, we further have the logical symbols '$\neg$', '$\Rightarrow$', and '$\forall$', corresponding—modulo the typed nature of the system—to usual negation, (material) implication, and universal quantification, respectively. As L(**IT**) is a language that is both substitutional and typological, one will find the symbolism characteristic of substitutional theories enabling one to express, for entities p, q, a, b, the proposition that q results from the replacement of a by b in p, in the form: $p(b/a)!q$. Beyond their particular use in formulas of that form, parentheses will serve in a general way for punctuation.

And now, here is L(**IT**).

1. Alphabet

   1.1 The variables 'x^n_m', when $n \geq 0$ and $m \geq 1$.

   1.2 Possibly, constants 'a^n_m', when $n \geq 0$ and $m \geq 1$.

   1.3 '$\neg$', '$\Rightarrow$', '$\forall$', '/', '!', '(', ')'.

2. Complete symbols, formulas, sentences

   2.1 The variables and the constants having n as a superscript are complete symbols of order n. The occurrence of a variable is real in the complete symbol that it constitutes.

   2.2 If $\mathbf{c}_1$ and $\mathbf{c}_2$ are complete symbols of orders $n_1 \geq 1$, $n_2 \geq 1$ respectively, then $\ulcorner(\neg\mathbf{c}_1)\urcorner$ and $\ulcorner(\mathbf{c}_1 \Rightarrow \mathbf{c}_2)\urcorner$ are complete symbols of orders $n_1, \max(n_1, n_2)$ respectively. The occurrences of variables that were real (apparent respectively) in $\mathbf{c}_1$ or $\mathbf{c}_2$ remain so within the new complete symbols.

   2.3 If $\mathbf{c}$ is a complete symbol of order $n \geq 1$, and $\mathbf{x}$ a variable of order $m \geq 0$ having a real occurrence in $\mathbf{c}$, then $\ulcorner\forall\mathbf{x}\mathbf{c}\urcorner$ is a complete symbol of order $\max(n, 1 + m)$. The occurrences of variables that were real (apparent respectively) in $\mathbf{c}$ remain so in the new complete symbol, except for the occurrences of $\mathbf{x}$ which were real, which are now apparent.

   2.4 If $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$, $\mathbf{c}_4$ are complete symbols of orders n_1, n_2, n_3, n_4, respectively, then $\ulcorner(\mathbf{c}_1(\mathbf{c}_2/\mathbf{c}_3)!\mathbf{c}_4)\urcorner$ is a complete symbol of order $\max(1, \max_{1\leq i\leq 4} n_i)$.[14] The occurrences of variables which were real (apparent respectively) in one of the symbols $\mathbf{c}_i$ remain so within the new complete symbol.

   2.5 The formulas are complete symbols of order ≥ 1.

   2.6 The sentences are the formulas without real occurrences of variables.

The system **IT** would be obtained from language L(**IT**) by adding axioms and rules of inference corresponding to its intended interpretation.

[14] If all the symbols $\mathbf{c}_i$ are of order 0, the order is thus equal to that which elementary propositions must have according to Russell, namely 1; if not, it is equal to $\max_{1\leq i\leq 4} n_i$.

6. The Intended Interpretation again

Russell kept almost completely silent about the language and, *a fortiori*, about the system of the intermediate theory. He was content to describe the hierarchy of individuals and propositions corresponding to its intended interpretation, after which he said a few words about the substitutional method in accordance with which the logical reconstruction of a hierarchy of functions would be possible on this basis. Finally, he abandoned the intermediate theory, pleading reasons of convenience, and moved on to the ramified theory of types—all this in barely three pages ([11]: 236–239, [18]: 75–77). The description of the hierarchy of individuals and propositions takes up a little bit more than a page.

1. I shall first quote Russell's text at length, furnishing it with some initial comments in square brackets. Let us remember that, despite appearances, the notions at issue in this text, in particular the notion of variable, are *extra*-linguistic notions.

> Propositions which contain apparent variables [more precisely: universally quantified variables: *vide infra*] are generated from such as do not contain these apparent variables by processes of which one is always the process of *generalization*, *i.e.* the substitution of a variable for one of the terms of a proposition [i.e. for one of its constituents in the subject position: *vide infra*], and the assertion of the resulting function for all possible values of the variable [the notion of function here serves to describe the hierarchy, but there will not be any function in this hierarchy]. Hence a proposition is called a *generalized* proposition when it contains an apparent variable. A proposition containing no apparent variable we will call an *elementary* proposition. (…) In an elementary proposition we can distinguish one or more *terms* from one or more *concepts*; the *terms* are whatever can be regarded as the *subject* of the proposition, while the concepts are the predicates or relations asserted of these terms. The terms of elementary propositions we will call *individuals*; these form the first or lowest type. [Although Russell does not say anything about it here, the concepts just in question may in turn be considered subjects in other elementary propositions, as terms of these propositions: they are themselves individuals.]
>
> By applying the process of generalization to individuals occurring [in subject position] in elementary propositions, we obtain new propositions. (…) We may define an individual as something destitute of complexity; (…) propositions are essentially complex. (…)
>
> Elementary propositions together with such as contain only [variables of] individuals as apparent variables we will call *first-order propositions*. These form the second logical type.
>
> We have a new totality [after that of individuals], that of *first-order propositions*. We can thus form new propositions in which [variables of] first-

> order propositions occur as apparent variables. These we will call *second-order propositions*; these form the third logical type. (…)
>
> The above process can be continued indefinitely. The $n+1$th logical type will consist of propositions of order n, which will be such as contain [variables of] propositions of order $n-1$, but of no higher order, as apparent variables. The types so obtained are mutually exclusive. ([11]: 237–238, [18]: 75–77)

2. Certain complete symbols of the language L(**IT**) do not seem to have any interpretation within the Russellian hierarchy, viz. the complete symbols introduced by rule 2.4. How is one to interpret them within this hierarchy? Indeed, the question could just as well be asked for the symbols introduced by rule 2.2, and the response would again be the same. Although Russell was not explicit on this point, we have to understand that propositions do not reduce to those generated by the iterated application of operations of *predication* and of *generalization*. (Here I conveniently call "predication" the operation that consists of joining a concept, predicate or relation, and an appropriate number of individuals given in a certain order to form a proposition.) To these operations one must add the usual operations of *negation* and (material) *implication*. But one also has to take into account a new operation, crucial to the intermediate theory, viz. *substitution*. Here I mean precisely the possibility of forming new propositions corresponding to the complete symbols introduced by rule 2.4. Russell was not as explicit as he should have been, and as I shall be in a moment, about the roles of negation, implication, and substitution in describing his hierarchy, but then is no hidden difficulty here.

Conversely, there is something in the Russellian hierarchy which seems not to be reflected in the syntax of the language L(**IT**). It is the structure of the propositions that result from the operation of predication alone: the so-called atomic propositions. However, there is nothing new here to be worried about. That was already the case with the first two substitutional theories: in the same sense—if not to the same degree—as they were, *the intermediate theory is a theory of unanalyzed propositions*.[15]

3. One will note the "constructive" nature, in a certain sense, of the Russellian hierarchy. Once the individuals are "given", divided up into particulars and universals (concepts), and the latter into predicates (qualities), two-place relations, three-place relations, etc., the totality of the hierarchy is "constructed" step by step in a unique way. The elementary propositions are propositions that one can "construct" out of individuals by iterated application of the operations of predication, negation, implication and substitution, *and nothing else*. The first-order propositions are obtained by adding the operation of generalization with the individual variables to the preceding operations, *and nothing more*. The idea of a first-order proposition that would be "given", so to speak, without being "constructed" out of individuals in the way indicated is ruled out. Likewise, the second-order propositions are obtained from first-order propositions by

[15] For a detailed analysis of this point with respect to the first substitutional theory, see [5]: chap. V, §8.

employing a certain procedure optionally involving negation, implication and generalization with variables of individuals, and necessarily involving generalization with variables of first-order propositions, *and nothing more*. And so on.

The kind of constructivism in question here must not be mistaken for other possible readings of this notion. I shall therefore distinguish:

1° the logical constructivism, or *constructivism*$_1$, that induced Russell, beginning in 1905 and 1906, to replace dubious entities (paradigmatically classes) with logical constructions (that is the idea of the *no-classes theory*), and which would find its radical expression, beginning in 1914, in the "supreme maxim of scientific philosophizing": "*Wherever possible, logical constructions are to be substituted for inferred entities*" ([14]: 388, [16]: 115);

2° the "constructivism" at issue here, or *constructivism*$_2$;

3° constructivism as a well-known doctrine of the ontological status of abstract objects, or *constructivism*$_3$; according to it these objects do not exist independently of ourselves, but are only mental constructions. This brand is classically contrasted with realism ("Platonism"), and always remained completely foreign to Russell.

In a nutshell, then, we can say that the hierarchy of types was constructive$_1$ in 1906, it became constructive$_2$ in 1908, but it would never be constructive$_3$.

There is no need to wonder why Russell limited himself to a constructive$_2$ hierarchy. He did not have to "limit" himself; rather, the very idea of proposition implied "limitation" right from the start. The idea of propositions that might not have their place within the hierarchy in question did not occur to Russell any more than the idea of natural numbers not finding their place in the intended ("standard") model of arithmetic had done to Frege, Dedekind, Peano or Russell himself.

4. I shall describe the Russellian hierarchy as I see it in more precise terms. Let us consider, independently of any requirement of effectiveness, the following language, L*(**IT**). Its definition broadly overlaps that of language L(**IT**) given in section 5. To define L*(**IT**), one provides oneself with a multiplicity of constants 'a_i^0' (i ranging over a certain multiplicity) in one-to-one correspondence with the individuals assumed to be "given" at the base of the hierarchy in the Russellian description, and for each of them a "degree": 0 if it corresponds to an individual that is not a concept; 1 if it corresponds to a predicate; $n \geq 2$ if it corresponds to a n-ary relation. At that point, the inductive definition of L*(**IT**) can begin. It will be enough to say that it *literally* coincides with the definition of L(**IT**) except for the following two amendments:

1° replace 1.2 with 1.2*:
1.2* The constants 'a_i^0' with their respective degrees $n \geq 0$;

2° complete 2.1 in 2.1*:
2.1* a) [repeat 2.1];
b) if **a** is a constant of degree n, and $\mathbf{b}_1$, …, $\mathbf{b}_n$, variables of order 0 or

constants (of any degree), then $\ulcorner(\mathbf{ab}_1 \ldots \mathbf{b}_n)\urcorner$ is a complete symbol of order 1; the occurrences of variables there are real.

One can then obtain a relatively precise idea of the Russellian hierarchy by describing it as the ontological transfer of the hierarchy of constants and sentences of L*(**IT**), except that one has to identify the propositions corresponding to the alphabetical variants of a single sentence of L*(**IT**) with one another. The absence of constants of order ≥ 1 in L*(**IT**) corresponds to the absence of any "given" entity that is not an individual in the hierarchy described by Russell—in other words, to the constructive$_2$ nature of that hierarchy. It will be understood that the sentences introduced by clause b) of rule 2.1* correspond to the propositions obtained from individuals by the operation of predication alone in the Russellian hierarchy. Language L(**IT**) does not itself contain any symbolism of this kind: the intermediate theory does not take the analysis that far, it does not penetrate the mystery of the ultimate structure of propositions.

7. Programmatic Note for the Development of IT

Russell abandoned the intermediate theory for the ramified theory of types for reasons of convenience. By way of conclusion, I shall say a word about the manner in which **IT** could be developed.

Let us first note that **IT** is, in its own way, a theory of types, but that—contrary to what generally happens in type theory—the differences of types do not correspond to any particular morphological taboo. These differences involve the interpretation of the symbols and would only leave a trace in the axioms and rules of inference.

From the morphological point of view, one could easily verify that the definitions given in the first substitutional theory for the logical construction (construction$_1$) of functions could be taken up again in L(**IT**) with appropriately adapted commentary.[16] By their successive definitions, functions of one, two, three, etc., (real) arguments would be arranged beyond individuals and propositions in accordance with a hierarchy of types. This hierarchy would be much more complex than the corresponding hierarchy of the first substitutional theory because the functions should be typologically distinguished according to the types of their arguments and the order of their values. One should, however, note that from the perspective of the substitutivity *salva congruitate* for the corresponding incomplete symbols, the hierarchy would actually be as simple as it was in the first substitutional theory, because this substitutivity would only depend on the number of arguments. This simplicity would be a consequence of the purely semantical nature of the determination of types for complete symbols. The classes of entities and the relations among entities would be defined in terms of these

[16] Like Russell from 1908 onward, one would no longer distinguish classes and relations in intension from the corresponding functions. By way of logical fictions, there would only be functions, on the one hand, and classes and relations in extension on the other.

functions, as Russell would do in the ramified theory of types. By their successive definitions, these classes and relations would be arranged in accordance with a hierarchy of types as simple as the corresponding hierarchy of the first substitutional theory. The other functions would be defined, as in the first substitutional theory, as particular cases of preceding ones, and the other classes and relations, in turn, in terms of these other functions.

As for the axioms and rules of inference, one would undoubtedly be led to a system comparable in strength to that of the system of the ramified theory of types, with the same problems,finally, and the same aporias.[17]

References

[1] Cocchiarella, Nino: 1980. The development of the theory of logical types and the notion of a logical subject in Russell's early philosophy. *Synthese* 45: 71–115.

[2] Hylton, Peter: 1980. Russell's substitutional theory. *Synthese* 45: 1–31.

[3] Landini, Gregory: 1998. *Russell's Hidden Substitutional Theory*. Oxford: Oxford University Press.

[4] Rouilhan, Philippe de: 1992. Russell and the vicious circle principle. *Philosophical Studies* 65: 169–182.

[5] Rouilhan, Philippe de: 1996. *Russell et le cercle des paradoxes*. Paris: Presses Universitaires de France.

[6] Rouilhan, Philippe de: 1996. Towards finishing off the axiom of reducibility. *Philosophia Scientiae* 1: 17–35.

[7] Russell, Bertrand: 1903. *The Principles of Mathematics*. London: Cambridge University Press. 2nd ed. G. Allen & Unwin, 1937.

[8] Russell, Bertrand: 1906. On the substitutional theory of classes and relations. Published in [19]. Paper read before the London Mathematical Society in 1906.

[9] Russell, Bertrand: 1906. Les paradoxes de la logique. *Revue de Metaphysique et de Morale* 14: 627–650. Original English version, On 'Insolubilia' and their Solution by Symbolic logic, published in [19].

[10] Russell, Bertrand: 1906–07. On the nature of truth. *Proceedings of the Aristotelian Society* 7: 28–49.

[11] Russell, Bertrand: 1908. Mathematical logic as based on the theory of types. *American Journal of Mathematics* 30: 222–262. Reprinted in [18].

[12] Russell, Bertrand: 1910 Cf. *Whitehead et Russell 1910–27* , vol. I, 1st ed.

[13] Russell, Bertrand: 1910. On the nature of truth and falsehood. In: *Philosophical Essays*, London: Longmans Green. 2nd ed. London: Allen & Unwin, 1966, 147–159.

[17] Thanks to Claire O. Hill for translating the penultimate version of this paper.

[14] Russell, Bertrand: 1914. The Relation of sense-data to physics. *Scientia* 4. Reprinted in [16]: 108–131.

[15] Russell, Bertrand: 1914. *Our Knowledge of the External World*. Chicago: Open Court. Rev. ed. London: G. Allen & Unwin, 1926.

[16] Russell, Bertrand: 1917. *Mysticism and Logic*. London: Allen & Unwin.

[17] Russell, Bertrand: 1918–1919. The philosophy of logical atomism. *The Monist* 28: 495–527; 29: 33–63, 190–222, 344–380. Lectures delivered in 1918; reprinted in [18]: 177–281.

[18] Russell, Bertrand: 1956. *Logic and Knowledge. Essays 1901–1950*. Ed. by R. C. Marsh. London: G. Allen & Unwin.

[19] Russell, Bertrand: 1973. *Essays in Analysis*. Ed. by D. Lackey. London: Allen & Unwin.

[20] Tarski, Alfred: *Projęcie prawdy w językach nauk dedukcyjnych*. Warsaw, 1933.

[21] Tarski, Alfred: 1936. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* 1: 261–405. Reprint dated 1935; English translation in [22]: 152–278.

[22] Tarski, Alfred: 1956. *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*. Trans. by J. H. Woodger. Oxford: Clarendon Press.

IHPST (CNRS and Université Paris I Panthéon-Sorbonne)
France

E-mail: rouilhan@ext.jussieu.fr

Propositional Ontology and Logical Atomism

Francisco Rodríguez-Consuegra

Abstract. In the following I will briefly indicate the role of propositional functions in *Principia*, then point out the way in which they continue to be introduced through propositions, in spite of the doubtful status of these pseudo-entities. I will then try to see if the notion of judgment, which is used by Russell to explain propositions, can meet the requirements which are needed in a serious, coherent ontology. After a survey of the different attempts to build up a convincing notion of proposition and judgment carried out by Russell between 1910 and 1918, I will conclude that the ontology of logical atomism was finally a failure, and so that the mathematical ontology usually associated with Russell's logicism was also a failure. In doing so, I will often show that almost all the problems in which Russell was involved in that period are somehow dependent upon what I will call Bradley's paradox of relations, according to which we cannot consider relations (or other similar "incomplete" notions) as genuine terms, or logical subjects. If we do that, then we should give an account of the way in which these relational terms are in turn related to other terms. But this leads to an unavoidable infinite regress.

'Logical atomism' is perhaps an appropriate expression for referring to the most important philosophy developed by Bertrand Russell, which is associated with his most important contributions to logic, foundations and philosophy of mathematics. These contributions were mainly presented in *Principia Mathematica* [23]. I will regard logical atomism as extending, broadly, from 1910 to 1918.

The ontology held by Russell in former periods of his philosophy (mainly in *The Principles of Mathematics* [7]) consists of terms (logical subjects), concepts (including properties and relations) and propositions, these last being genuine entities, possibly to be regarded as a special sort of terms. Here, classes are terms, and are used to define numbers and the rest of mathematical entities. The status of propositional functions should, on this scheme, be equivalent to that of concepts, and they, despite their being genuine primitive notions, could easily be introduced in terms of (albeit not reduced to) propositions, given that a propositional function, when a value is assigned to its variable, is transformed into a proposition (as its final value).[1] The problem in logical atomism, beginning in *Principia* or even earlier, is that propositions (in the sense that they are expressed by sentences) are no longer regarded as genuine entities, but merely as sets of entities, so the symbols used to denote them are no longer true symbols, but incomplete symbols. So, propositional functions can hardly be explained through propositions, unless we first clarify the ontological status of propositions.

[1] For the details of the whole construction in *Principles*, see my [5].

In *Principia* propositional functions, together with the usual quantifiers, are used to define-eliminate: descriptions; classes, through the axiom of reducibility (which is also named "axiom of classes"); relations, through the corresponding axiom of reducibility (or "axiom of relations"), and also the fundamental notions of identity and membership. From the viewpoint of mathematical ontology, that means that although numbers, the basic notion on which mathematics is founded, are defined as classes of classes. And, since classes are nothing but their members and everything that you may want to say about classes can be said through propositional functions, numbers, as well as the rest of the mathematical entities, can legitimately be reduced to propositional functions. Yet, propositional functions are nothing but properties, so numbers are ultimately properties.

Let us first examine how Russell describes the relationship between propositional functions and propositions: "By a 'propositional function' we mean something which contains a variable x, and expresses a *proposition* as soon as a value is assigned to x" ([23]: vol. I, 38). So, propositional functions cannot even be understood without appealing to the notion of proposition. When you assign a value to the variable in a propositional function ($\phi\hat{x}$) you obtain, as the value of the function, a proposition (ϕx). Since propositional functions are genuine entities, and since the values you assign to their variables are entities too (values which belong to the class $\hat{x}(\phi \mathrm{x})$), you accordingly have to obtain entities as final values. That is, you can have sentences as values of propositional functions if these are just expressions; you can have propositions as values if propositional functions are treated as entities, but you cannot have propositional functions as entities and sentences as values. Therefore, propositions should be true entities, and according to Russell, they are denoted by propositional functions. Although denotation was interpreted merely as a relation between words and entities (not as a logical relation between concepts, as it was in *Principles*), propositions should be seen as "objects", as Russell writes:

> The function itself, $\phi\hat{x}$, is the single thing which ambiguously denotes its many values; while ϕx, where x is not specified, is one of the denoted objects, with the ambiguity belonging to the manner of denoting. ([23]: vol. I, 40)

Yet the problem starts when Russell writes in *Principia* that propositions are not "single" entities, but sets of them which are in need of judgment to reach the necessary unity, judgment being a "multiple" relation. This theory of judgment, which was held by Russell in unpublished manuscripts as early as 1906 or 1907, see my [4]. However, it was not admitted in public until the official proclamation in 1910 ([23]: vol. I, 41–44) ([10]: 155–8[2]), where judgment is openly explained as a multiple relation between the mind and several terms which, taken together, constitute a complex as a consequence of the judgment itself:

[2]Curiously enough, these pages are part of the new essay that Russell wrote to replace the third section of [8] (which was here eliminated with no explanation), where he still presented the multiple relation theory only as one more possibility.

> Owing to the plurality of the objects of a single judgment, it follows that what we call a "proposition" (in the sense in which is distinguished from the phrase expressing it) is not a single entity at all. That is to say, the phrase which expresses a proposition is what we call an "incomplete" symbol; it does not have meaning in itself, but requires some supplementation in order to acquire a complete meaning. This fact is somewhat concealed by the circumstance that judgment in itself supplies a sufficient supplement, and that judgment in itself makes no *verbal* addition to the proposition. Thus "the proposition "Socrates is human"" uses "Socrates is human" in a way which requires a supplement of some kind before it acquires a complete meaning; but when I judge "Socrates is human," the meaning is completed by the act of judging, and we no longer have an incomplete symbol. ([23]: vol. I, 44)

Thus, propositions are not single entities, but rather complexes of entities, which are joined through a special relation. However, propositions (sentences) are now regarded as incomplete symbols, for they need the complement of the judgment itself to reach full meaning, and it is unclear whether this suggests that propositions as complexes are, in fact, nothing at all. In principle, the complex, being a mere creation of the mind, would be something subjective, and this would be a sign that Russell was also trying to avoid one of the forms of Bradley's paradox concerning the statement of any form of correspondence.[3]

The problem lies in explaining how it is possible that the (judging) mind is not a part of the complex which is supposed to be the object of the judgment, while it is, at the same time, to be taken as a part of the multiple relation of judging itself. For in this way the mind can impose a certain order among the terms with which it is acquainted, an order which does not exist on its own. Thus, the status of relations as terms, i.e., as genuine realities, is doubtful, given that, although we accede (through acquaintance) objectively to the constituents of the complex, yet the relation joining them is a mental product (the judging relation). And this, again, turns into the problem about the pre-eminence of relations over terms or vice versa.

On the other hand, it is truly surprising that, for Russell, an incomplete symbol (the proposition before being judged by the mind) can be transformed into a complete one (the proposition once judged by the mind); for an incomplete symbol designates nothing, while a complete one designates something extra-linguistical. So the mind seems to have the incredible capacity to transform nothing into something.

Russell's continued his defense in 1907 of the objective status of external relations as the only correct alternative to monism, while he was really espousing, in other manuscripts, the infeasibility of treating propositions as complexes, as well as the subjectivity of the judging relation, is, in my view, incoherent. As usual, he had to find a compromise between two strong forces pulling in opposite directions. On the one

[3] I.e., the *relation* of correspondence between two things is in need of another relation which in turn relates it to the two things, and so on.

hand, he needed to point out relations as terms[4] to escape from the image he constructed about monism. On the other, he also needed to deny the reality of these relations as terms in one of the most important instances (that of propositions and judgments) in order to avoid Bradley's paradox (now under the form of the impossibility of regarding complexes as terms), while, at the same time, he should avoid any correspondence theory of truth as falling down into another form of the same paradox. Yet some form of such a correspondence was necessary to avoid idealism. But the inconsistencies caused by these two forces appear even more clearly when we consider the details of the multiple relation theory as stated by Russell in *Philosophical Essays* [10] (as in *Principia* he avoids any treatment of such problems).

Let us see how Russell poses the problem of false judgments. In general, the "natural" view concerning judgment says that there is a relation (judging) between our mind and a certain object, the judgment being true or false according to the truth and falsehood of the object of judgment. But Russell had two objections against false objects ([10]: 175ff). First, they seem to depend on the full significance that the corresponding sentence acquires through the addition of the expression "I believe that . . .". This is the line leading to judgments as "incomplete symbols". Second, if we admit false objects, we are admitting entities whose existence does not depend upon the existence of judgments, which seems to Russell something incredible, as "it leaves the difference between truth and falsehood quite inexplicable".

Thus, on the one hand, Russell wants to accept some sort of a corresponding entity whose presence or absence can determine the truth or falsehood of the judgment, while, on the other hand, he wants to deny that the absence of this corresponding entity may mean the "existence" of a false object. However, we cannot say that true judgments have objectives while false ones do not, "so long as we hold the view that judgment is actually a relation of the mind to an objective. For . . . a relation cannot be a relation to nothing" ([10]: 176–7). The only solution seems dividing the object of judgment into several parts:

> The way out of the difficulty consists in maintaining that, whether we judge truly or whether we judge falsely, there is no one thing that we are judging. When we judge that Charles I died on the scaffold, we have before us, not one object, but several objects, namely, Charles I and dying and the scaffold . . . . We therefore escape the necessity of admitting objective falsehoods, or of admitting that in judging falsely we have nothing before the mind. ([10]: 177)

Therefore, the "unity of the judgment" cannot be provided by a relation joining the different elements present in it to form a complex, with which the mind is then uniformly related. Rather, that unity is provided by the relation between the mind and each of those elements (the terms and the original relation), "i.e., we must have, not several instances of a relation between two terms, but one instance of a relation between more

[4]Recall that Russell uses the word *term* for *entity* in its most general sense, cf. [7]: 43.

than two terms". This was the solution Russell discovered to the Bradleyan problem of reconstructing complexes.

The later official formulation is this:

> Every judgment is a relation of a mind to several objects, one of which is a relation; the judgment is *true* when the relation which is one of the objects relates the other objects, otherwise it is false . . . . Let us take the judgment 'A loves B'. This consists of a relation of the person judging to A and love and B, i.e., to the two terms A and B and the relation 'love' . . . . The 'corresponding complex' object which is required to make our judgment true consists of A related to B by the relation which was before us in our judgment. ([10]: 181, 183)

In this way, we can certainly avoid false objects, while some form of correspondence is maintained, although obviously at a rather high price. For, on the one hand, the unity of the judgment seems to depend upon some sort of mental operation which is in charge of ultimately joining the terms and the relation in the judgment; but, on the other, the new theory ignores the ontological advantages of not distinguishing between judgment and perception in the way of Moore,[5] which is involved in Moore's rejection of the correspondence theory of truth. Last, but not least, there is also the problem of the "sense" of the relation between the different terms in the judgment, which has to be related, in some way, to the mind. One could get the impression that Russell, in discussing these problems, was somehow uncertain of his new theory.

Russell's doubts concern precisely the most difficult point: the relation between the two complexes involved, i.e., the relation between the two relations which give these complexes their unity.[6] Russell writes: "judgment is a relation of the mind to several other terms: when these other terms have *inter se* a 'corresponding' relation the judgment is true; when not, it is false" ([10]: 153). A first sign of Russell's doubts is the fact that the relevant correspondence appears with inverted commas. A second sign is the fact that the problem of the "sense" of the involved relations is avoided, while Russell says only that this sense has to be the "appropriate" one:

> We may distinguish two 'senses' of a relation according as it goes from A to B or from B to A. Then the relation as it enters into the judgment

[5] Concerning this last point, Russell says:

> One of the merits of the above theory is that it explains the difference between judgment and perception, and the reason why perception is not liable to error as judgment is …. Thus in perception I perceive a single complex object, while in a judgment based upon the perception I have the parts of the complex object separately though simultaneously before me. ([10]: 181–2)

However, the old Moorean realism made it impossible to make a distinction between fact, or complex object, i.e., perception, and judgment. Thus, any perception involved a judgment (in the sense of a proposition), which made it impossible to maintain any theory of correspondence.

[6] That is, r_1(mind, A, r_2, B) being the main complex, and where $r_2(A, B)$ is the secondary one, how are r_1 and r_2 related to one another?

> must have a 'sense', and in the corresponding complex it must have the same 'sense'. Thus the judgment that two terms have a certain relation R is a relation of the mind to the two terms and the relation R with the appropriate sense: the 'corresponding' complex consists of the two terms related by the relation R with the same sense. The judgment is true when there is such a complex, and false when there is not. ([10]: 183–4)

This point is very important, for it involves precisely the connection between the two relations, and as one of these relations is mental and the other objective, the danger of idealism seems undeniable. Perhaps such an unsolved problem was the reason why Russell said that his theory preserved the necessary "mixture of dependence upon mind and independence of mind" ([10]: 158).

Later in the same work, Russell stresses the same point by saying that the new theory succeeds in preserving truth and falsehood as properties of judgments. Thus, they are properties which, in a certain sense, depend upon the existence of minds, while, on the other hand, "the truth or falsehood of a given judgment does not depend upon the person making it or the time when it is made, since the 'corresponding' complex, upon which its truth or falsehood depends, does not contain the person judging as a constituent" ([10]: 184). Yet I cannot help feeling that in this way Russell comes very close to some sort of contradiction: truth and falsehood of judgments depend upon the mind and then again do not depend upon it. This might be one of the consequences of having forgotten Moore's radical way of avoiding the problem, which consisted in rejecting any correspondence theory by identifying judgment with perception.

In *Problems of Philosophy* [15] Russell already explicitly considered the problem of the sense of the relations involved, but provided no solution to it. The starting point is already misleading, for in pointing out several examples of multiple relations, he refers only to multiple relations among people, while the example which is actually analyzed later is one in which one of the terms is a relation: "Othello believes that Desdemona loves Cassio". Russell then admits two different senses, according to the two relations involved, but he continues to maintain that the judging relation provides its sense to the secondary relation: "the relation 'loving', as it occurs in the act of believing, is one of the objects—it is a brick in the structure, not the cement. The cement is the relation 'believing'" ([15]: 128). However, this can only mean that the terms are put in a certain order precisely "by the 'sense' of the believing" "(*ibid.*), with which, if they have no order by themselves, it is difficult to see how it is possible to speak of the 'correspondence' between the two complexes.[7]

[7] There is another indication of the extreme importance of the problem of sense in Russell's correspondence with Broad. See for instance this letter to Russell of June 25, 1912:

> Suppose we have an n-adic relation. Then the number of possible [?] complexes of the same terms will be $|\underline{n}$. In a judgement about such a complex we shall have $n + 2$ terms, i.e. [?] the n terms of the original relation, the relation itself, and the judging mind. So that believe will involve a $n + 2$-adic relation and therefore $|\underline{n + 2}$ psychical states differing only in the sense of their relating relation will be possible. If for a true belief each complex

Moreover, the problem concerning idealism remains unaltered, as does the problem concerning Bradley's paradox, in both forms of the endless regress: clarifying the relation between relation and terms, and the complex indicating the correspondence between the two previous complexes. If the judging relation imposes a sense on the secondary complex, then this sense is a product of the mind, given that the mind is one of the constituents of the judging relation. On the other hand, if the mind has acquaintance with each member of the complex,[8] then it has to be acquainted with the *relation* as well, and it seems that this should include the involved sense. If not, not only does the judging relation impose its own sense on the secondary complex (with which we might produce "nonsense", as Wittgenstein later objected), but also we must explain how it is possible to *connect* this second relation to the related terms, with the subsequent danger of arriving at Bradley's regress in the first form.[9]

> of the object terms must be correlated with one and only one of the psychical complexes then ought to be for any given complex of objects $|\underline{n+2} - 1$ possible false beliefs as to the sense. But actually there are only $|\underline{n} - 1$.
>
> Would you say that $|\underline{n+2} - |\underline{n}$ of the possible psychical complexes are not beliefs, or that as a matter of fact they never exist? If so on what principle do you decide what are beliefs or which exist? And, apart from this, can we tell [?] at all which sense of the object complex is correlated with which one of the belief complex? Finally the relation of belief in different beliefs has different polyadicity according to that of the relation of the object complex. Is there [?] any other example of the same relation having different polyadicity? I think you hold that relations do not have instances; what then is there in common between all the differently polyadic relations that relate the complexes which are all called beliefs?

And Russell's response of Jan 31, 1912:

> It is not the case that $|\underline{n+2}$ psychical states differing only in the *sense* of the relation of judging are possible. A mind must occupy one fixed position in the complex. E.g., "B. judges that the sun is shining" is possible, but not "the sun judges that B. is shining". If several of the other constituents of the judgments are minds, of course the case is altered [?]. But this is only possible when God is the Judge, since we have no acquaintance with other minds. And then polytheism is required really. Generally, your principle that $|\underline{n}$ complexes can be made of n terms is wrong; many senses will not yield a complex if the terms are of different sorts.
>
> The question whether we can tell which sense of the object-complex is correlated with which of the belief-complex is more difficult. It is plain to me that we can, but I hardly know how. I think it must come by way of acquaintance with a perceived complex. You perceive "A-to-right-of-B" and you judge that A is to right of B; this gives a correlation of the two senses. Thence it could be extended to cases when the complex is not perceived. But I am not sure that this is satisfactory.
>
> The question of the different polyadicity of different relations of belief is difficult too. I have been always inclined to suppose these [?] were different relations of belief, 3-term, 4-term, etc. But then, as you say, one wants to know what they all have in common, and to that I don't know what to answer. Perhaps only an associated feeling. The question is serious and I should be glad to know the answer.

Both letters are extant in the Russell Archives, McMaster University, Hamilton, Canada.

[8] In [15] Russell openly admits this point: "when we are judging, we have *a* relation to each of the constituents of our judgment separately" ([15]: 153), but he avoids the term "acquaintance" and replaces it by "being conscious of", perhaps to make the problem seem less important.

[9] Anyway, it is also unclear how we are to understand Russell when he says that the sense is ultimately provided by the judging relation, for the real problem in jealousy does not, I am afraid, lie in whether

The second form is unavoidably implied by the supposed 'correspondence' between the two complexes. We must not forget, first, that for Russell, at this stage, there are no propositions *per se*, for they (the corresponding sentences) are mere incomplete symbols which exist only in judgment (or beliefs, which are synonymous for Russell in this context). Thus there is no complex ArB on its own—there is only "I belief that ArB." At this point the possibility of the correspondence is quite obscure, for as it is the judging relation that imposes its sense, thus actually producing a complex (which does not exist independently), it is hard to accept that there must be a correspondence between the judging relation, which is a product of the mind, and the resulting secondary complex, which is also a product of the mind, if the theory of incomplete symbols is to be accepted. The endless regress is present here, just as it is in any other form of correspondence: to explain the relation of correspondence between the first complex, which is a belief (the judging relation), and the second complex (the secondary relation), we need another belief stating the fact of the correspondence, as Russell wrote many times when he was attacking the correspondence theory at former stages of his philosophy.

However, here the second kind of endless regress can be presented even under another more subtle form, which was already suggested by Stout in a further criticism [22], in spite of the fact that Stout did not see any problem of the form of a Bradleyan endless regress. The fact is, just as there must be a correspondence between the two complexes, we also have to know whether or not the judgment-complex is itself apprehended, for we must *compare* it with the secondary complex. But "this would imply that whenever we believe we must at the same time be aware of the state or process of believing, and of the mind as a constituent of it" ([22]: 343). In other words, if we have "I believe that Ar_1B", and we need to compare the two complexes "r_2(Mind, A, r_1, B)" and "Ar_1B" to know about the possible correspondence and then about the truth of the judging relation, then we also need "I believe that $\{r_2(\text{Mind}, A, r_1, B)\}\ r_3\ \{Ar_1B\}$", and so on.

The climax of all these problems can be found in the articles in which Russell considered the status of relations and predicates, in an attempt to set up a whole ontology in which a full classification of universals and particulars was provided, once the multiple relation theory had been publicly admitted. In those articles we can see how the inconsistencies between the two forces we have described above come close to being full-fledged contradictions. In the following, I shall point out some of these inconsistencies in the four main papers published between 1911 and 1912.

In [11] we first find the argument that we have "awareness" of universals, and especially of relations, although the "proof" is not very convincing: "yellow differs from blue". But treating relations as things that are immediately known to us supposes that their recognition has an objective ontological status which, although derived ul-

---

Desdemona loves Cassio or Cassio loves Desdemona, but rather in whether or not Desdemona loves Cassio. And this indicates that the mind can not only provide a sense, but also produce a relation which does not exist *as relating Desdemona and Cassio*. (Although perhaps this is not quite the same as saying that the mind creates the relation at all, for Russell believed that we have acquaintance with universals.)

timately from Russell's need for regarding them as genuine terms (then as ultimate realities), is hardly compatible with *Principia* where relations, like classes, are regarded as mere fictions, and their names rendered as incomplete symbols accordingly. It can be said that the relations which were considered there were only relations in extension, but to say that is not to say much, for classes had the same extensional status, and Russell never tried to admit them as objective realities "in intension" in his ontology, while they had to be rejected in logic, which is extensional (except, of course, through propositional functions, which, being ultimately properties, are intensional entities).

This would be a constant ambiguity in the rest of Russell's development, and I think it ultimately proceeds from the tension between the two forces we have been considering since the beginning of this paper. The only argument we are given is that we not only directly know complexes that contain relations as constituents, but also relations as logical subjects, which is a consequence of the principle of acquaintance, according to which any proposition we can understand must be composed of constituents with which we are acquainted ([11]: 117). However, the principle presupposes that there is already a solution to the problem of relations as terms, which, as we have seen, can hardly be maintained.

Another sign of Russell's problems with relations regarded as "forms" involves the difficult status of propositional functions themselves. In *Principia,* as we have seen, they were to be regarded as primitive ideas to which all the rest of our "formal" concepts (i.e., classes, relations, descriptions, etc.) could be reduced. Here he adds that propositional functions are "complexes" which play the role of "true subjects" or "ultimate subjects" ([11]: 126, 128), despite the fact that this would suppose (i) that we would have acquaintance with them, and (ii) that they would be true constituents of complexes, as they appear in judging multiple relations. Of course Russell needed to regard propositional functions as logical subjects in *Principia*, but he provided no theory to explain these two consequences. This, together with the problems pointed out above, might have contributed to the changes in the multiple relation theory of judgment in *Theory of Knowledge*,[10] where "forms" are openly admitted as being parts of the multiple relation, as we shall see below.

The essay [12] tries to build up a whole philosophical view, epistemological and ontological at the same time, but the problems of relations as terms remain unsolved. Russell starts by writing:

> It [my philosophy] is analytic for it maintains that the existence of what is complex depends upon the existence of what is simple, and not vice versa, and that a constituent of one complex is absolutely identical, as constituent, to what it is in itself when its relations are not considered. ([12]: 53)

However, Russell seems to forget that relations are also constituents of complexes, so that some explanation of their status as simples must be provided, especially because

[10] Written in 1913, but not published until 1984. See below.

it is unclear how it is possible that relations are exactly the same as the rest of the constituents of a complex, i.e., how it is possible to regard relations both as relating relations and as mere constituents that are not related at all. The problem again has to do with the twofold nature of relations: the paradoxical consequences we have been seeing force us to recognize them simultaneously as terms and concepts.

That is why when Russell classifies the different kinds of "beings" in the world, he is silent about the place where relations are to be found: "I say then that there are simple beings in the universe, and that these beings have relations through which they compose complex beings" ([12]: 56). He admits that every complex has two kinds of constituents: terms and relations (or predicates). So he seems to have forgotten that, according to his own view, every complex must also contain relations *as terms*, so the distinction is difficult to accept. Finally, since Russell admits predication as a true universal, he should provide some explanation of the ontological status of predication, and even of the consequences concerning the old theory of judgment, according to which there is no ontological difference at all between subject and predicate. But nothing of the kind is to be found in this paper.

The paper [13] tries to provide some responses to these problems. We have to remember that Russell needed, at the same time, to consider predicates as relations (or as involving relations, which give their true essence) and also as predicates in themselves. He needed the former, because in this way some form of the old Moorean theory of judgment could still be maintained (and so the danger of the predicative general form avoided); but he needed the latter too, because propositional functions, the most important entities in *Principia*, are nothing but the general form of properties (and therefore they seem to be able to reduce all propositions to the predicative form).

In this paper Russell introduces the difference between monadic and dyadic concepts, doubtless in an attempt to face the difficulty, but is then forced to speak of predicates as having the two different natures: "It is of course the case that, whenever a subject has a predicate, there is a dyadic relation of subject and predicate, but it does not follow that there is not also a proposition in which the predicate is merely predicated" ([13]: 159). That implies further that the analogy with relations is complete: we have predicates as relations ("predicating" predicates) and predicates as terms (predicates "in themselves"), which leads exactly to the same unsolved problems as with relations (for we also have "relating" relations along with relations "in themselves"). The first form avoids Bradley's paradox but cannot provide terms, and the second form provides true terms but cannot explain complexes unless we introduce further relating constituents and accept Bradley's paradox. The following passage is a further effort to make sense of the distinction:

> Whenever a has the relation R to b, there is a triadic relation of a and R and b, but in this relation R occurs as a term of the relation, not as the relating relation of the proposition. Similarly, if there are monadic concepts, the proposition in which they are said to have the relation of predication to their subjects will not be identical with the propositions in which they are actually predicated. ([13]: 159)

But the new manoeuvre is only the old strategy already used to avoid Bradley's paradox, for here we have two additional relations which do not appear in the notation: one is the triadic relation which actually relates the rest of the constituents; the other is the relation *R* no longer as a relation, but as a term. Thus, to explain a dyadic relation we need a triadic one where the former relation acquires a different status; so presumably to explain the triadic relation we are going to need a quadruple one, and so on. Since the same goes for predicates, we get the same problem Russell encountered in former stages: there is no point in asking whether or not predicates can be transformed into relations, for predicates are already relations.

The paper [14] can be regarded as the final stage of the series. Its main goal was to defend the ultimate dualism between universals and particulars, which Russell needed in order to separate concepts (predicates and relations) from terms. But he also needed concepts to be terms as well, which made any consistent solution impossible.[11] Thus, the classification itself is vitiated by the old problem of the twofold role. That is why when Russell says that particulars enter into complexes only "as the subjects of predicates or the terms of relations", and universals "as predicates or relations" ([14]: 124), he seems to be oblivious of the fact that, according to his own position in other contexts, predicates can be subjects and relations can be terms.

Here Russell openly admits that predication is a relation "involving a fundamental logical difference between its two terms" ([14]: 123), but this is not much, for at the same time he needs to admit the twofold role of predicates, which makes the supposed logical difference impossible. As Russell himself confesses, he needs to maintain a specific relation of predication to be able to make an ultimate distinction between particulars (those which cannot be predicates or relations) and universals (those which are only predicates or relations). But as we have seen the distinction cannot be ultimately maintained, for it amounts to nearly the same thing as saying, in a language close to Bradley's, that universals are already impregnated with particularity. At this stage Russell was no longer able to accept such contradictory metaphors.

I do not know how far these efforts can be located in relation to Bradley's theory that universals are also particulars and vice versa, but, at any rate, it seems that the corresponding idealism is somehow involved in Russell's view. We will also discover that the monism (and the holism) implicit in this idealism will appear more and more openly in later stages.

Anyway, the polemic which Russell maintained with Bradley directly, both in the publications and in the unpublished correspondence, can be usefully considered in this context. Bradley started the (published) polemic in [1], by accusing Russell of maintaining an inconsistent pluralism, for he admitted "unities which are complex and which cannot be analysed into terms and relations" ([1]: 176). Russell's response in [9] tried to avoid the ultimate inconsistency by saying that he did not maintain that unities are incapable of analysis, but only that the mere enumeration of the constituents cannot

[11] In *Principles* Russell juggled with the problem by introducing a further division of terms, into things and concepts. But then the problem remains when we have to give an account of the fact that relations—concepts in principle—can also be regarded as things.

reconstruct the unity: "A complex differs from the mere aggregate of its constituents, since it is one, not many, and the relation which is one of its constituents enters into it as an actually relating relation, and not merely as one member of an aggregate" ([9]: 344). But in saying this he provides no true reply to the objection, as Bradley lucidly pointed out in [2]:

> Is there anything, I ask, in a unity beside its "constituents", i.e., the terms and the relation, and, if there is anything more, in what does this "more" consist? Mr. Russell tells us that we have got merely an enumeration or merely an aggregate. Even with merely so much I should still have to ask how even so much is possible. But, since we seem to have something beyond either, the puzzle grows worse.

The personal correspondence[12] shows Russell to be much more appreciative of Bradley's philosophy. In a letter from 1907, Bradley had already pointed out that wholes can by no means be reconstructed, for they are non-relational in the last analysis: "But if you break this entity up, and set down any part as independent—then, starting with this part, there is no getting beyond it except arbitrarily …. You will say that you replace this by external relations. But it is denied that these serve" (Oct. 21, 1907). Russell admitted the difficulties, but only by claiming that relatedness does not imply complexity, with no further explanation of how we can give an account of relations as *related* to their terms (Oct. 29, 1907).

The final stage of the correspondence is captured very nicely in this passage by Russell, which contains the failed attempt to escape the old paradox: "I do not consider pluralism incompatible with the existence of complex entities. I consider that in every case where two simples have a relation, there is a complex entity consisting of the two simples so related" (April 9, 1910). But this, again, misses the point, for the expression "complex entity" involves different unexplained senses. We only have to replace "simples" with "terms" to realize that Russell is forgetting that the relation is also a term on his own view. This in turn requires him to explain the difference between the complex as formed by the three terms, and the complex as being the "complex entity", supposedly formed by the terms as "simples" and the "relating relation". This is why Russell finally added a third entity, the "form", and faced the same paradox at a higher level.

Russell's final attitude implied the recognition that even considering the recent publications from 1910–11, the problem of unities remained unsolved: "I have nothing short to say, the subject is difficult (…), and I do not pretend to have solved all its problems" (March 2, 1911). Fortunately, there is still another passage showing the strong link between Bradley's criticisms, Russell's unsolved difficulties, the new theory of judgment, and Wittgenstein's objections (Jan. 30, 1914):

> I fully recognise the vital importance of the questions you raise, particularly as regards "unities"; I recognise that it is my duty to answer if I can,

[12] In the Russell Archives, MacMaster University, Hamilton, Canada. A more detailed study of this correspondence will appear in my *Relational Ontology and Analytic Philosophy,* now near completion.

> and, if I cannot, to look for an answer as long as I live . . . Chiefly through the work of an Austrian pupil of mine, I seem now to see answers about unities; but the subject is so difficult and fundamental that I still hesitate.

To my knowledge, this is the only place where Russell admitted these important links, as well as the fact that he regarded his former views as a failure, precisely from the viewpoint of Bradley's objections, and not only as regards Wittgenstein's.

The next stage, which might perhaps be entitled: "Enter Wittgenstein: the multiple relation theory staggers on", can, I think, be regarded, to some extent, as the time at which Russell's philosophy had to pay the price for his continuous delay in facing the "fundamental principles" (as Bradley used to say). This price turned out to be quite high, for it involved the abandonment of a major project in the philosophy of logic and epistemology, and his subsequent devotion to particular problems, without any hope of finding an acceptable global philosophy. However, the difficulties which made the project impossible were already present in former stages, as already pointed out by Bradley, despite the fact that Wittgenstein's criticisms were given the credit for Russell's disappointment with his own philosophy.[13]

The manuscript [16] is already an attempt to characterize the notion of form. But the starting point returns to former views: "The *form* of a complex is what it has in common with a complex obtained by replacing each constituent of the complex by something different".[14] Thus, we have to avoid Bradley's paradox, for if we make the form a constituent, "it would have to be somehow related to the other constituents, and the way in which it was related would really be the form; hence an endless regress" ([16]: 2). Therefore, though no final definition of the notion is provided, Bradley's paradox is respected, and for the same reason any possible violation of the theory of types was apparently avoided, for any attempt to consider forms on the same ontological level as constituents would be such a violation, for it would regard "formal" concepts as individuals or substances.

That is why it is so strange that in the unfinished book of 1913, *Theory of Knowledge* [17], Russell's main recourse was the admission of forms as constituents of complexes with no explanation of the type-theoretical problems involved. Of course, this was also an attempt to avoid the criticisms of Bradley and Wittgenstein, which forced him to introduce some changes to avoid the rather idealist consequences of regarding complexes, and then propositions, as a mere creation of our minds. But these criticisms forced Russell to face a very unpleasant dilemma: he could either maintain the old theory of types by abandoning the multiple relation theory of judgment and by renouncing the attempt to characterize complexes, and then logic, or abandon some philosophical consequences of the theory of types by complementing the multiple theory with an explicit device making the admission of forms as some sort of constituent possible. Russell's solution in 1913 was clearly a compromise between the two horns

[13] Griffin's [3] also contains a good survey of those criticisms.

[14] The examples which Russell mentions are propositional functions, dyadic and multiple relations, and the two standard forms of quantification.

of the dilemma, although the final solution seems to involve the second alternative, despite Bradley's prohibition.

Russell now rejected the view that the form can be a "mere" constituent of complexes with the usual Bradleyan argument. Hence, in "Socrates is human," "is" represents the form, and thus cannot be a constituent: "for if it were, there would have to be a new way in which it and the two other constituents are put together, and if we take this way as again a constituent, we find ourselves embarked on an endless regress" ([17]: 16). But when he tries to explain the nature of form, he rejects regarding it not only as an equivalence relation (i.e., "to be the same form"), but also as a mere primitive idea; for it would lead us to the usual paradox, when trying to relate this idea with the others in the system. The solution, already found in Wittgenstein's framework, was to regard it as an indefinable object corresponding to certain general expressions.

Thus, the form of subject-predicate complexes will be "something has some predicate," and the form of dyadic complexes will be "something has some relation to something." Russell tries to avoid the obvious attack of circularity by writing: "in spite of the difficulties of language, it seems not paradoxical to say that, in order to understand a proposition which states that x has the relation R to y, we must understand what is meant by 'something having some relation to something'" ([17]: 114). However, the compromise supposes only one change in the general scheme of the already published multiple relation theory of judgment, viz., incorporating a symbol of the form (γ) into the general complex constituting the judgment: "$U(S,\ x,\ R,\ y,\ \gamma)$". Thus, it is difficult to deny that forms are regarded as constituents. However, this is not only an attempt to leap over Bradley's paradox, but also to give an objective status to the form as something with which we are acquainted, in the same way as a relation was both a relating relation and a term. The only change is now to regard γ as the general form of dual complexes, and to show it to absorb the "relating" part of R, though, of course, we still need to explain the status of γ in the complex.

I cannot enter here into Wittgenstein's criticisms in any detail[15] which forced Russell to leave the book unpublished, but for Wittgenstein the main defect of Russell's revised theory was that all the difficulties involved had their common root in the attempt to regard the form as a new constituent. Wittgenstein expressed this in a language shrouded in mystical connotations, but this cannot cover the fact that he clearly saw the impossibility of Russell's attempt to fix the nature of a form without simultaneously using another form. Thus, Bradley's paradox is also present: we cannot meaningfully *speak* about the relation between the judging complex and the secondary complex.

If we look at his publications at this stage, Russell, rather surprisingly, did not abandon the multiple relation theory of judgment, nor renounce the need to regard forms as constituents. In [18] Russell starts by declaring forms to be the object of "philosophical logic" ([18]: 52), although in resorting to the "replacement" device, he says only that "form is not another constituent, but is the way the constituents are put together". Thus, without mentioning his failure to construct an acceptable

[15] I did it in [6].

epistemology of logic, he even adds the explicit claim that we have knowledge of forms, allowing us to understand sentences: "Thus some kind of knowledge of logical forms, though with most people it is not explicit, is involved in all understanding of discourse" ([18]: 53). We have thus the two traditional arguments, but nothing about the status of the form in the judging complex, while the multiple theory is apparently maintained, for Russell denies objective negative facts except as false beliefs: "It is therefore necessary, in analysing a belief, to look for some other logical form than a two-term relation" ([18]: 66). In the face of that statement, I am unable to see any weight in the usual claim that Russell "abandoned" the multiple relation theory in 1913 because of Wittgenstein's criticisms.[16] On the whole, then, I must conclude that Russell made no progress in trying to solve the real problems underlying all of his rather edifying talk of relations.

The last paper [19] was Russell's final attempt to maintain a consistent theory before officially abandoning the multiple relation theory of judgment. Yet, we again find exactly the same unsolved problems. As regards forms, the same idea of constituting an inventory is introduced by merely changing "forms of propositions" to "forms of facts" ([19]: 216), which surfaces again as a "realistic bias" despite the need for providing an account of false "facts" in terms of some kind of multiple relation theory. This multiple theory seems *to still be maintained here*, for it is said that belief is not a dual relation: "Therefore the belief does not really contain a proposition as a constituent but only contains the constituents of the proposition as constituents" ([19]: 224).[17] Bradley's paradox is presented a few pages later, when Russell admits that we cannot put the verb on the same level as its terms because it is an instance of form, which can by no means be a further constituent: "the form of the dual relation … is not a constituent of the proposition. If it were you would have to have that

[16] Russell seems to have maintained some sort of multiple relation theory in his lectures in America. The following are some notes from V. Lenzen's "Notes on Russell lectures" (now extant in the Russell Archives), which were taken in March 1914, that is to say, after Russell had left *Theory of Knowledge* unfinished and, supposedly, abandoned the theory of judgment it contained. In Lenzen's notes we can read, for example, about propositional attitudes as involving not a dual but a multiple relation:

> Judgment: all objects must be things with which you are acquainted …. Acquaintance with universal—logical form of occurrence—not same as acq. with particulars. Possibility of error in any cognitive occurrence shows that occurrence is not dual relation …. I believe Jones hates Smith—single fact—contains 2 verbs. Constitutes oddity of propositional thought …. Logical form of occurrence is different from that of presentation.

Lenzen's term paper dealt precisely with Russell's theory of judgment, and it contained two main criticisms: judgment cannot be a relation, for (i) truth *is* a relation; that is why we say there are true judgments; (ii) relations are universals, while judgment is a process in time. Russell's reply, in the form of notes added to the paper, reads: "*Judgment* is a relation, *a judgment* is not a relation. Thus *man* is a universal, but *a man* is not. Your argument (…) on this point sins against philosophical grammar". Also: "'A judgment' will be a positive fact in which the principal relation is *judging*; but a judgment is not itself a relation. What is related to an objective is judgment, not a judgment". Thus, Russell does not confess his strong doubts concerning the multiple relation theory after Wittgenstein's Bradleyan criticisms, but he seems to continue to maintain the theory rather *explicitly*.

[17] For, as stated before, "Every fact that occurs in the world must be composed entirely of constituents that there are, and not of constituents that there are not" ([19]: 220).

constituent related to the other constituents" ([19]: 239). The problem is, then, the same: Bradley's paradox makes the form as a constituent impossible, but precisely this is required by the multiple relation theory.

The persistence of these ideas can even be seen once again in [20], written in 1918 ([20]: 198 ff). Here Russell needed some definition of logic,[18] for the book was explicitly devoted to the logicist foundations of mathematics. He resorts to the usual "solution": "we may accept, as a first approximation, the view that *forms* are what enter into logical propositions as their constituents," with which "logic is concerned only with *forms*, and is concerned with them only in the way of stating that they are always or sometimes true." Thus, he says not only that forms are constituents, but also that we can have second-order propositions in which we refer to these forms and the rest of the constituents. He is obviously forgetting the two forms of Bradley's paradox. However, a few lines earlier he wrote: "the form is not itself a new constituent; if it were, we should need a new form to embrace both it and the other constituents."

If I have to propose some conclusion here, I can only add that the chaotic philosophy usually called "logical atomism", can hardly be regarded as "logical" or as "atomism". It is not logical, for it was mainly the compromise between two incompatible, purely philosophical (ontological) traditions: that of Moore, and that of Bradley. Neither is it atomistic, for it was mainly devoted to including forms in the ultimate inventory of the world, which was impossible as it was attempted by making forms to have a twofold nature, that of terms, and that of non-terms. Furthermore, this philosophy was hardly realistic, for the unavoidable idealistic implications of the multiple relation theory of judgment, together with the monism implicit in the logical constructions, made it impossible to continue to maintain that complexes are genuine objective realities. They are, at best, the result of *our* application of forms to the terms we perceive in the world. But since these terms, even when they are the simplest possible, can be further constructed out of relational structures, which are ultimately nothing but "incomplete symbols", the old realistic and atomistic world is replaced by an inescapably idealistic and holistic one, where terms can no longer be opposed to relations, where relations can no longer be regarded as true terms, and where the seed of the later abandonment of the distinction between subject and object was already present.

Acknowledgement. I am grateful to an anonymous referee for useful suggestions.

[18] In a letter of 1918—to Frank Russell—Russell says that the most important things still to be reached at that stage were: (i) a theory of judgment; (ii) a definition of logic. The new orientation, as to the *solutions*, is now openly psychological, but the *problem* is still the same: the unity of complexes. I think it is worth quoting the letter:

> "Facts, Judgments, and Propositions" opens out—it was for its sake that I wanted to study behaviourism, because the first problem is to have a tenable theory of judgment. I see my way to a really big piece of work, and incidentally to a definition of "logic", hitherto lacking. All the psychology that I have been reading and meaning to read was for the sake of logic; but I have reached a point in logic where I need theories of (a) judgment (b) symbolism, both of which are psychological problems. ([21]: 249)

References

[1] Bradley, Francis H.: 1910. On appearance, error and contradiction. *Mind* 19: 154–185.

[2] Bradley, Francis H.: 1911. Reply to Mr. Russell's explanations. *Mind,* 20, 74–76.

[3] Griffin, Nicholas: 1986. Wittgenstein's criticisms of Russell's theory of judgment. *Russell* 5: 132–145.

[4] Rodríguez-Consuegra, Francisco: 1989. Russell's theory of types, 1901-1910: its complex origins in the unpublished manuscripts. *History and Philosophy of Logic* 10: 131–164.

[5] Rodríguez-Consuegra, Francisco: 1991. *The Mathematical Philosophy of Bertrand Russell*. Basel: Birkhäuser. Reprinted in 1993.

[6] Rodríguez-Consuegra, Francisco: 2002. Wittgenstein y la teoría russelliana de la proposición. In my *Estudios de filosofia del lenguaje*, Granada: Comares, 179–214. Partial english translation of this paper, with some changes and new material, as "Wittgenstein and Russell on propositions and forms". In J. Padilla Gálvez (ed.), *Wittgenstein from a New Point of View*. Wittgenstein-Studien, vol. 6, Frankfurt a. M.: Peter Lang, 2003, 79–110.

[7] Russell, Bertrand: 1903. *The Principles of Mathematics*, Cambridge: The University Press. London: Allen & Unwin. Second edition in 1937. Paperback edition London: Routledge, 1992.

[8] Russell, Bertrand: 1907. On the nature of truth. *Proc. Arist. Soc.* 7: 28–49.

[9] Russell, Bertrand: 1910. Some explanations in reply to Mr. Bradley. *Mind* 19: 373–378.

[10] Russell, Bertrand: 1910. *Philosophical Essays*. London: Longmans Green.

[11] Russell, Bertrand: 1911. Knowledge by acquaintance and knowledge by description. *Proc. Arist. Soc.* 7: 108–128.

[12] Russell, Bertrand: 1911. Le réalisme analytique. *Bull. Soc. Franç. Phil.* 11: 53–82.

[13] Russell, Bertrand: 1911. The basis of realism. *Jrn. Phil.* 8: 158–161.

[14] Russell, Bertrand: 1912. On the relation of universals and particulars. *Proc. Arist. Soc.* 12: 1–24.

[15] Russell, Bertrand: 1912. *Problems of Philosophy*. London: Williams and Norgate.

[16] Russell, Bertrand: 1912. What is logic? In: *The Collected Papers of Bertrand Russell,* vol. 6, London: Routledge, 1992.

[17] Russell, Bertrand: 1913. *Theory of Knowledge*: the 1913 manuscript. In: *The Collected Papers of Bertrand Russell,* vol. 7, London: Allen & Unwin, 1984.

[18] Russell, Bertrand: 1914. *Our Knowledge of the External World*. London: Allen & Unwin.

[19] Russell, Bertrand: 1918–1919. The philosophy of logical atomism. *The Monist* 28–29. Repr. in *Logic and Knowledge*. Edited by R. Ch. Marsh, London: Allen & Unwin, 1956.

[20] Russell, Bertrand: 1919. *Introduction to Mathematical Philosophy*. London: Allen & Unwin, 198ff.

[21] Russell, Bertrand: 1919–1926. *The Collected Papers of Bertrand Russell,* vol. 9. London: Unwin Hyman, 1988.

[22] Stout, Georg F.: 1915. Mr. Russell's theory of judgement. *Proc. Arist. Soc.* 15: 332–352.

[23] Whitehead, Alfred N., and Bertrand Russell: 1910, 1912, and 1913. *Principia Mathematica*, 3 vol's. Cambridge: Cambridge University Press.

Departament de Lògica i Filosofia de la Ciència
Facultat de Filosofia i CC de l'Educació
Avinguda Blasco Ibáñez, 30 - 7ª
46010 València
Spain

E-mail: Francisco.Rodriguez@uv.es

Classes of Classes and Classes of Functions in *Principia Mathematica*

Bernard Linsky

Abstract. The No-Classes theory of classes in *Principia Mathematica* must be supplemented with a definition of propositional functions of classes in order to avoid a recently discovered puzzle. Such a definition is suggested but rejected by Whitehead and Russell in the Introduction to *PM* and does not make it into the technical presentation of ∗20. *PM* thus contains a distinction between classes of classes and classes of functions although only the former appear in the body of the work. The problem that justifies this distinction is distinct from a problem about scope identified by Carnap in 1947 in *Meaning and Necessity*, although Carnap's problem also involves the contextual definitions of the no-classes theory and the intensional nature of propositional functions.

The "No-Classes Theory of Classes" in ∗20 of *Principia Mathematica* provides what Russell had ultimately come to accept as the solution to his paradox of the class of all classes that are not members of themselves. It provides definitions of the extensional notions of the theory of classes within a logic of intensional propositional functions. While the theory of classes is conducted in what amounts to a simple theory of types, the underlying logic of propositional functions is the "ramified" theory of types. The ramification, justified as it is by its connection with the vicious circle principle which bans impredicative definitions, seems to be associated with a constructivist attitude towards functions, and hence, presumably, towards classes as well. The excursion into ramification of the types of propositional functions is apparently then undone by adopting the axiom of reducibility. Or at least this has been one standard view of the role of ramification and reducibility in *PM*.[1] In recent years some students of *PM* have followed Alonzo Church in promoting an "intensional interpretation" of that work, reasserting the significant role of intensional notions including that of propositional functions in its logical foundations.[2] If the fundamental nature of the logic of *PM* is to be a ramified type theory of intensional propositional functions, then whatever its independent justifiability, the axiom of reducibility becomes at least

[1] See [9]: 35. Surprisingly, from Church himself, ([2]: 355), we find a criticism of the Axiom of Reducibility: "Indeed, as many have urged, the true choice would seem to be between the simple functional calculi and the ramified functional calculi without axioms of reducibility. It is hard to think of a point of view from which the intermediate position represented by the ramified functional calculi with axioms of reducibility would appear to be significant."

[2] Including L. Linsky [7] and Church [2].

understandable for its necessary role in reconstructing the extensional theory of classes within that larger intensional logic. Whatever element of constructivism there might be in the theory of propositional functions does not extend to the theory of classes that is reproduced within the larger logic.

This paper proposes to extend the intensional interpretation of *PM* by showing how Whitehead and Russell distinguish between classes of classes and classes of propositional functions. Failing to observe that distinction leads to what will look like a serious problem for the no-classes theory. Part of the worry about the axiom of reducibility, and indeed the role of functions in the work is that once classes are introduced propositional functions seem to disappear from *PM*.[3] The vast bulk of the project of reducing mathematical notions to logic in fact consists of reducing mathematical objects to classes of classes with appropriate features. Not all higher order classes in the system of *PM* are classes of classes, however. Admittedly the notion of classes of functions plays no significant role in the body of the work. Indeed Whitehead and Russell remark "We shall never have occasion explicitly to consider classes of functions, but classes of classes will occur constantly …"([*PM*]: 190). Nevertheless, the apparatus to make the distinction is discoverable in the details of ∗20 and particularly in an obscure passage near the end of the Introduction where a definition is proposed but then seemingly rejected. This paper will present the case for this distinction, motivating the need for it, and attempt to clarify that obscure passage. Along the way, the details of the system which are able to control the interaction of intensional phenomena with extensional classes will be presented. The result will be to show the extent to which the logic of *PM*, even when it concerns classes, is thoroughly intensional.

The heart of the so called "no-classes" theory of ∗20 is a list of contextual definitions of all occurrences of class expressions and variables ranging over classes which is inspired by the famous contextual definition of definite descriptions in ∗14. That theory allows definite descriptions to occur in all syntactic contexts in sentences where a singular term could occur, and then provides an analysis of those contexts yielding the correct truth conditions for those sentences. It is important to notice that contextual definitions of so called "incomplete symbols", if correctly carried out, allow for the elimination of those symbols from *all* contexts in which they might appear. There is no role for ill formed expressions using incomplete symbols beyond those that derive from the type restrictions on propositional functions in the underlying logic. Consequently the theory of classes will have to account for expressions asserting that a function is in a class, namely a class of functions, but also for classes being in those classes of functions. As well there will be classes as members of what are clearly classes of classes, but an analysis is provided for assertions that functions belong to such classes.

The definitions of ∗20 have the effect of reducing talk about classes that are extensional, in that two with the same members are identical, to expressions about

[3]This observation is made by Quine ([9]: 251).

propositional functions which are "intensional", in that two distinct functions can be true of the same arguments. The principal definition achieves this by saying that a function will be true of a class, just in case it is true of some function coextensive with the defining function of that class. In the notation of *PM* this is:

$$*20{\cdot}01 \quad f\{\hat{z}(\psi z)\}. =: (\exists \phi) : \phi!x \,.\equiv_x .\psi x : f\{\phi!\hat{z}\} \quad Df$$

To be more precise, the function f is true of the class of ψs just in case some predicative function ϕ (as indicated by the !) which is coextensive with ψ has f. The role of the notion of predicative function and the accompanying axiom of reducibility complicates the interpretation of the no-classes theory, but is independent of the problem discussed here. That problem would arise for any attempt to define classes in this sort of way within a theory of intensional properties.

One of the primary contexts in which class expressions will occur, and which needs to be defined, is in conjunction with the membership sign ϵ. In first order axiomatizations of set theory, this is taken as the one non-logical primitive. Once talk of classes is interpreted as talk about propositional functions, the definition of this notion is straightforward; an object is *in* a function just in case the function is true of it:

$$*20{\cdot}02 \quad x \,\epsilon\, (\phi!\hat{z}). = .\phi!x \quad Df$$

$*20\cdot02$ should be read as "typically ambiguous", that is, while the individual variable x is used in the number, and so presumably $\phi!$ will be a function of a type that applies to individuals, the whole should be seen as schematic, allowing for an interpretation which says what it is for a function of individuals f to be "in" a function of such functions ([*PM*]: 189). If $\phi!$ is such a higher order function of functions of a given type, then $\hat{z}\phi!z$ will be a class of functions of that type. (Classes of functions will be written $\hat{f}\Phi!f$ below, with f a function of a given type and Φ a function of a type that can apply to f)

To complete the elimination of class expressions from all contexts, definitions must be provided to allow for the interpretation of occurrences of bound variables that range over classes. *PM* uses the lower case Greek letter α for such variables. Roughly, what is true of some class will be what is true of some function which has that class as extension. More precisely:

$$*20{\cdot}07 \quad (\alpha).f\alpha. = .(\phi).f\{\hat{z}(\phi!z)\} \quad Df$$

$$*20{\cdot}071 \quad (\exists\alpha).f\alpha. = .(\exists\phi).f\{\hat{z}(\phi!z)\} \quad Df$$

Class variables may also appear in class abstracts, that is, in expressions for classes of classes. Those occurrences are defined this way:

$$*20{\cdot}08 \quad f\{\hat{\alpha}(\psi\alpha)\}. =: (\exists \phi) : \psi\alpha. \equiv_\alpha .\phi!\alpha : f\{\phi!\hat{\alpha}\} \quad Df$$

With these definitions the properties of classes that Whitehead and Russell need for the rest of *PM* can be derived, with only the (notorious) addition of axioms of reducibility, infinity and choice (called the "multiplicative" axiom). The effect is to capture the power of set theory done in the simple theory of types.[4]

One of the most fundamental results of the no-classes theory is to allow the proof as a theorem what otherwise is adopted as the "axiom" of extensionality:

$$*20{\cdot}31 \vdash :. \hat{z}(\psi z) = \hat{z}(\chi z) \,.\equiv:\, x \in \hat{z}(\psi z) \,.\equiv_x .\, x \in \hat{z}(\chi z)$$

Again, observing the "typical ambiguity" of the theory, this theorem will apply to higher order classes of functions as well. No corresponding theorem is proved saying that classes of classes are identical if they have the same classes as members. I will argue below that such a theorem for classes of classes is not already asserted by this schema, but could be stated and proved.

This list of contextual definitions given in $*20$ would seem to be incomplete, however. While uses of Greek letters as variables bound by universal and existential quantifiers can be eliminated, as well as those bound by class abstracts, as in $\hat{\alpha}(\psi\alpha)$, no definition is given for free occurrences of Greek letters, and none of occurrences of Greek letters in expressions for functions of classes, as in $\phi!\hat{\alpha}$ in $*20{\cdot}08$, with the exception of:

$$*20{\cdot}081 \quad \alpha \in \psi!\hat{\alpha} .= \psi!\alpha \quad Df$$

In fact there are definitions for these other occurrences suggested in the *Introduction* to *PM* at page 80, which are curiously not included in the text of $*20$. The main argument of this paper will result in an explanation of these tentative definitions, providing motivation for their inclusion and an account of their oddity: an oddity which seems to have prevented Whitehead and Russell from including them in the body of the text.

As a comment on $*20{\cdot}42$ which uses a free occurrence of α that should read as having universal force, we have "A Greek letter, such as α, is merely an abbreviation for an expression of the form $\hat{z}(\phi z)$ …" ([*PM*]: 194). Although there is no formal definition of propositional functions with Greek variables in the body of $*20$, we do find this obscure "definition" tucked in almost at the very end of the Introduction ([*PM*]: 80-81):

> Accordingly
>
> $$f\alpha = .(\exists\psi).\phi!x \equiv_x \psi!x.f\{\psi!\hat{z}\} \qquad Df.$$
>
> It is a peculiarity of a definition of the use of a single letter [viz., α] for a variable incomplete symbol that it, though in a sense a real variable, occurs only in the *definiendum*, while "ϕ," though a real variable, occurs only in the *definiens*.

[4] See ([5]: Chp. 4, §4.3) for the system **TT** of simple type theory, and the following discussion.

Thus "$f\hat{\alpha}$" stands for

$$\text{“}(\exists\psi).\hat{\phi}!x \equiv_x \psi!x.f\{\psi!\hat{z})\text{"}$$

and "$(\alpha).f\alpha$" stands for

$$\text{“}(\phi) : (\exists\psi).\phi!x \equiv_x \psi!x.f\{\psi!\hat{z})\text{"}$$

This is indeed a peculiar definition, but at least certain cases of it can be clearly stated, using the language of *PM*, and the need for such a definition will be clear in what follows.

We can approach an understanding of this passage by considering how, according to the definitions presented in ∗20, one would analyse a sentence asserting that a class of individuals is in a second level class of classes, say, $\hat{x}\,f\,x \in \hat{\alpha}\Phi\alpha$. The following apparent derivation using the definitions above leads to a proper analysis at 5) but is missing two crucial additional steps.

1) $\hat{x}\,f\,x \in \hat{\alpha}\Phi\alpha$

2) $(\exists g)(\ g!x \equiv_x f\,x\ \&\ g \in \hat{\alpha}\Phi\alpha)$ using ∗20·01

3) $(\exists g)(\ g!x \equiv_x f\,x\ \&\ (\exists\Psi)(\ \Psi!\alpha \equiv_\alpha \Phi\alpha\ \&\ g \in \Psi!\hat{\alpha}))$ by ∗20·08

4) $(\exists g)(\ g!x \equiv_x f\,x\ \&\ (\exists\Psi)(\ \Psi!\hat{x}h!x \equiv_h \Phi\hat{x}h!x\ \&\ g \in \Psi!\hat{\alpha}))$ by ∗20·07

5) $(\exists g)(\ g!x \equiv_x f\,x\ \&\ (\exists\Psi)(\ \Psi!\hat{x}h!x \equiv_h \Phi\hat{x}h!x\ \&\ \Psi!g))$ applying ∗20·02?

Expressions of the sort $g \in \Psi!\hat{\alpha}$ require careful attention. $g \in \Psi!\hat{\alpha}$ does not mean simply $\Psi!g$. More than a simple substitution of one for the other is required to move from 4) to 5). What follows is a digression through the problems of Martin and Carnap which show the need for the alternative account of expressions like $\Psi!\hat{\alpha}$ and the accompanying need for a distinction between classes of functions and classes of classes such as $\hat{\alpha}\Phi\alpha$.

One might think, following Gregory Landini ([6]: 169), that one could analyze 1) by first eliminating the class term on the right, namely $\hat{\alpha}\Phi\alpha$, taking the context $\hat{x}\,f\,x \in \ldots$ in which it occurs taken as the f in ∗20·08. Then using ∗20·081, the right result is obtained and every step is provided for in the definitions of ∗20. Landini in fact argues that the other order of elimination requires the very sort of analysis that I will propose below, but rejects the needed definition as unacceptable, relying simply on Russell's reservations about its "peculiarity". One reason, however, for pressing on with the left to right order of elimination of the class terms is Whitehead and Russell's statement that "With regard to the scope of $\hat{z}(\psi z)$, and to the order of elimination of two such expressions, we shall adopt the same conventions as were explained in ∗14 for $(\iota x)(\phi x)$" ([*PM*]: 188). The discussion of conventions on scope in ∗14 includes this: "Hence we shall in general adopt the convention that the description occurring first typographically is to have the larger scope, unless the contrary is explicitly indicated"

([*PM*]: 174). That this particular part of the policy on scope is also intended for classes is confirmed by a remark in ∗20: "The meaning of '$\hat{z}(\psi z) = \hat{z}(\chi z)$' is obtained by a double application of ∗20·01 to ∗13·01 [the definition of identity], remembering the convention that $\hat{z}(\psi z)$ is to have a larger scope than $\hat{z}(\chi z)$ because it occurs first"([*PM*]: 191). So it appears that following the conventions of *PM* one will confront contexts of the form $g \in \Psi\hat{\alpha}$ in the analysis of any simple formula which asserts that one class is in another. We must look elsewhere for an account of membership in classes of classes.

A key to the understanding of expressions such as $\Psi!\hat{\alpha}$ is supplied by finding the solution to an apparent problem for the no-classes theory recently identified by D. A. Martin. He suggests that the no-classes theory of ∗20 produces "too many" classes.[5] An example of these seemingly excess classes if provided by considering two non-identical, but empty functions, i.e., functions true of nothing. It seems that there would then be at least two distinct second order classes whose only member class is the empty class. As Russell and Whitehead identify the class containing only the empty class with zero, this particular example may also be seen as showing that there are "too many zeros" in *PM*.

Here are details of this example of Martin's problem. Assume that there are two (predicative) functions of the type that apply to individuals, say f_1 and f_2, that are distinct, yet are both true of nothing, so:

6) $\sim (f_1 = f_2)$

and

7) $\sim \exists x f_1 x \ \& \sim \exists x f_2 x$

Suppose further that there are two second level functions, Φ_1 and Φ_2 which are uniquely true of f_1 and f_2 respectively:

8) $(f)(\Phi_1 f \equiv f = f_1) \ \&(f)(\Phi_2 f \equiv f = f_2)$

By assumption 7), the class of f_1's is just the empty set (Λ) and so is the class of f_2's:

9) $\hat{x} f_1 x \ = \ \hat{x} f_2 x \ = \ \Lambda$

According to the 6) and 8) above, and by the definition of ϵ at ∗20·02, f_1 will be in the class of functions which are Φ_1 but not in the class of functions which are Φ_2 (letting f be a variable ranging over functions):

10) $f_1 \in \hat{f} \Phi_1 f \ \& \ \sim (f_1 \in \hat{f} \Phi_2 f)$

We now derive the surprising result that two classes of functions can have all the same classes as members, since both classes contain only the empty class Λ, yet still be distinct, since one has a function in it that the other does not. Since any function g coextensive with some function in $\hat{f} \Phi_1 f$ and hence with f_1, will also be coextensive some function in $\hat{f} \Phi_2 f$, namely f_2, we will have:

[5] As a comment from the audience during a presentation on the no-classes theory that I was making to the Philosophy of Mathematics Workshop, Department of Philosophy, UCLA, Nov. 29, 2000.

11) $[(\exists h)(h!x \equiv_x gx) \ \& \ h \in \hat{f}\Phi_1 f] \ \equiv_g \ [(\exists k)(k!x \equiv_x gx) \ \& \ k \in \hat{f}\Phi_2 f] \ \& \ \sim (\hat{f}\Phi_1 f \ = \ \hat{f}\Phi_2 f)$

and so, using $*20{\cdot}01$ twice,

12) $(\hat{x}gx \in \hat{f}\Phi_1 f \ \equiv_g \ \hat{x}gx \in \hat{f}\Phi_2 f) \ \& \ \sim (\hat{f}\Phi_1 f \ = \ \hat{f}\Phi_2 f)$

a sentence that surely shows that classes may be meaningfully said to be members of classes of functions. $*20{\cdot}01$ gives a precise meaning to such expressions. Then, applying $*20{\cdot}07$ to introduce the bound Greek variable β which ranges over classes, we have:

13) $(\beta \in \hat{f}\Phi_1 f \ \equiv_\beta \ \beta \in \hat{f}\Phi_2 f) \ \& \ \sim (\hat{f}\Phi_1 f \ = \ \hat{f}\Phi_2 f)$

Finally, using $*20 \cdot 071$ we seem to have a counter example to the "axiom" of extensionality (taking Γ and Δ to be Greek letters for higher order classes):

14) $(\exists\Gamma)(\exists\Delta)(\beta \in \Gamma \ \equiv_\beta \ \beta \in \Delta \ \& \ \sim (\Gamma \ = \ \Delta))$

The class which contains only the empty set is, in fact, defined to be the number 0 using the function $\iota`x$ which yields the singleton class containing x:

$*54{\cdot}01 \quad 0 \ = \ \iota`\Lambda \quad Df$

Several results are proved about 0, including:

$*54{\cdot}102 \quad \vdash : \ \alpha \in 0 \ . \equiv . \ \alpha = \Lambda$

If this definition is to be proper, and the results about 0 legitimately provable, it must be the case that there is only one class containing Λ. Our example above suggests that there are "too many 0's" in *PM*.

The consequence 14) of our suppositions, however, does not in fact contradict the theorem of extensionality $*20{\cdot}31$. If one distinguishes classes of classes and classes of functions, these distinct classes $\hat{f}\Phi_1 f$ and $\hat{f}\Phi_2 f$ will be seen to only agree on all *classes* which belong to them, not all functions. Hence they are not genuinely coextensive. One has a function as member that the other does not. If one did not distinguish classes of classes and classes of functions, one would have to identify all second order classes which agree on their member classes. The examples above would indeed show that there were too many classes in *Principia Mathematica*. Similarly, if one makes the distinction, while there will be many classes of functions that contain only the empty class, there will only be one class of classes to which it belongs, only one 0.

While this issue of too many classes arises from a combination of intensional functions with the extensional theory of classes, it is important to see that this is not simply a matter to be resolved with careful attention to the scope of incomplete symbols for classes, as is a problem that Rudolf Carnap raised for the no-classes theory. In his [1] Carnap pointed out that the combination of intensional predicates with the no-classes theory gives what are at the very least "awkward" results:

> … let us consider two non-extensional properties Φ_1 and Φ_2 such that Φ_2 is the contradictory of Φ_1; hence Φ_2 holds in all cases, and only in those, in which Φ_1 does not hold. Since Φ_1 is nonextensional there are different, but equivalent, properties, say f_1 and f_2, such that Φ_1 holds for f_1, but not f_2, and hence Φ_2 holds for f_2. Then according to definition 33.2 [Carnap's version of $*20{\cdot}01$] both Φ_1 and Φ_2 hold for the class $\hat{z}(f_1 z)$, although Φ_1 and Φ_2 are contradictions and hence logically incompatible. This would be an awkward result, although it does not constitute a formal contradiction, since Φ_1 and Φ_2 are logically exclusive only with respect to properties …
> ([1]: 148)

My instances for Φ_1, Φ_2, f_1 and f_2 above do fit the specifications of this problem. The difficulty that Carnap sees, however, is that Φ_1 and Φ_2, while incompatible with respect to all properties, are both true of a common class, $\hat{z}(f_1 z)$. He does not say that $\hat{f}\Phi_1 f$ and $\hat{f}\Phi_2 f$ are then distinct classes of classes.

Although Carnap's own example for this problem described above might look very like the new problem Martin has identified, it is nevertheless different. Carnap considers the distinct but supposedly coextensive properties $f_1 = H\hat{z}$, $\hat{z}$ is a human, and $f_2 = F\hat{z} \bullet B\hat{z}$, $\hat{z}$ is featherless and a biped. For $\Phi_1 f$ he uses $f = H\hat{z}$, the property of being identical with the property of being human, and for Φ_2, $f \neq H\hat{z}$, *not* being identical to the property of being human. The "awkward" result we get is that $\hat{z}(Hz) = H\hat{z}$ and $\hat{z}(Hz) \neq H\hat{z}$.[6] This is not the problem Martin has identified. In our example what seems to be proved is that something is a member of one class and not the other, so they are distinct classes. The classes do not differ simply by there being some function co-extensive with the defining of one class that is not identical with something co-extensive with the defining function of the other class. Martin's problem cannot be dismissed as simply the result of forgetting that class abstracts need to be eliminated with care from any intensional context, including identity statements.[7]

If one follows Russell in taking class expressions to have narrow scope with respect to negation signs, then $\hat{z}(Hz) \neq H\hat{z}$ will not be true. What it says, according to this convention, is that *no* function coextensive with $H\hat{z}$ is identical with $H\hat{z}$, and that of course, is false. This response to Carnap's problem, combined with my response to Martin's problem, suggests that careful attention to scope conventions and the details of the contextual definitions of $*20$ resolves any awkwardness resulting from the combination of intensional functions and extensional classes.[8]

---

[6] Both this last example, of an identity between a class and a function, and the analysis of this in terms of scope seems to be originally due to Leon Chwistek ([3]: 14–16). Chwistek presents the example as part of an argument against the blanket use of the convention which Whitehead and Russell propose of always reading the scope of an incomplete symbol such as a description or class term to be as narrow as possible.

[7] See the Appendix to [7] for a discussion of this problem of Carnap's and a proposal for a solution by means of conventions about the elimination of class expressions.

[8] Gregory Landini ([6] §6.6 and 6.7), however, claims that Carnap has identified a pseudo-problem on the grounds that there is no possibility of scope ambiguity for class terms in these contexts. Definitions

Martin has also proposed a solution to his own problem.[9] The idea is to redefine ϵ, so that the members of a function, and hence of a class, will be just those properties which ultimately depend only on coextensive functions. The motivation for this approach is to see the no-classes theory as taking functions as proxies for classes, and then redefining the membership relation to eliminate the effects of the intensionality of functions. Martin's solution requires the simultaneous inductive definitions 15)–17) of a new equivalence relation $\approx$ and a redefinition of ϵ as $\in$:

15) $\quad f \in \phi \equiv \exists g\ (g \approx f\ \&\ \phi g),$

where f is one type lower than ϕ, and may be an individual variable.

16) $x^0 \approx y^0\ \equiv\ x = y$

For individuals x and y, of type 0, $\approx$ is just identity. Finally,

17) $g^{n+1} \approx f^{n+1}\ \equiv\ (x^n)(x \in g\ \equiv\ x \in f)$

where x is of any type one below f and g, as indicated by the superscripts.

Russell and Whitehead do not make use of any such definitions by induction on type, but using the device of typical ambiguity, they could easily be stated in *PM* notation. This definition blocks the derivation of the undesirable conclusion above, for the classes $\hat{f}\Phi_1 f$ and $\hat{f}\Phi_2 f$ will now have the same members, as determined by the new notion $\in$. In this approach all higher order classes are classes of classes. One need not distinguish classes of functions from classes of classes.

One more change must be made if the other theorems of set theory are still to be derivable using the new notion of $\in$ as representing membership. The "axiom" of extensionality as expressed in terms of $\in$ is not forthcoming on the definitions we have so far:

18) $x \in \hat{z}(\psi z).\ \equiv_x\ x \in \hat{z}(\chi z).\ \equiv:\ \hat{z}(\psi z) = \hat{z}(\chi z)$

In the example above, $\hat{f}\Phi_1 f$ and $\hat{f}\Phi_2 f$ are coextensive in this sense, though not identical.

The necessary alteration is to change the original clause of the no-classes theory, $*20{\cdot}01$ as follows:

19) $\quad f\hat{z}(\psi z)\ =_{df}\ \exists\phi(\ \phi \approx \psi\ \&\ f\phi!\hat{z})$

---

which are stated for singular terms and functions can only be applied after class terms are eliminated, and that can only be done after Greek letters, which he reads as schematic, are replaced by specific class terms. Whether or not a strict and unique order of elimination can be recovered from the various definitions in *PM*, Landini is only criticizing the particular case of $\neq$ as giving rise to issues of the order of elimination of derived terms. The more general description of the problem that Carnap gives at the beginning of the quoted passage presents problems about scope in intensional contexts without involving defined expressions other than those for classes.

[9] In an email to D. Kaplan, Nov. 30, 2000 [personal communication].

Using 19) as an additional clause of the revised definition of classes, 18) becomes provable. This completes Martin's suggested remedy and would allow for a reconstruction of the no-classes theory with theorems about classes and membership that coincides with the simple theory of types for classes.

This solution is not in accordance with the rest of *PM* ∗20, however. It comes at the cost of making all higher level classes be classes of classes. Thus note that it will no longer be the case that f_1 is *not* in the class of things that are Φ_1, for some function coextensive with it, namely f_2, surely is. The alternative solution proposed below keeps f_1 out of the class of functions that are Φ_2, but lets it into the class of classes that are Φ_2, not the only intuitive result, but one which at least allows for a distinction between the two sorts of higher level classes.

The solution I propose follows from paying closer attention to Russell and Whitehead's use of special Greek variables (such as α) to range over classes and makes sense of the "peculiar" definition in the Introduction. It also relies on following the model of the theory of descriptions very closely in observing that the no-classes theory does not take classes to be particular functions, but simply presents contextual definitions that are intended to allow proof as theorems of all the theory of classes that one needs. This includes expressions that talk about classes of classes, some using quantifiers, others with free occurrences of class variables. Describing the contents of ∗20 Russell and Whitehead say that "Lastly, we have a set of propositions designed to prove that classes of classes have all the same formal properties as classes of individuals" ([*PM*]: 189). They then go on to make the remark above about the lack of classes of functions in what follows. They could have added that their definitions also show how classes of functions have all the same formal properties as classes of individuals. Classes of functions are in the system, just never used in the body of *PM* and so not discussed.

The key is to see that $f\hat{\alpha}$ is defined in that peculiar passage in the Introduction as a function of functions, different from $f\hat{\phi}$. We should take the suggested definition in the Introduction as proposing that $\Phi\hat{\alpha}$ be defined as $\Phi\hat{x}\,\hat{f}x$, which in more modern notation would be written $(\lambda f)(\Phi\{x : fx\})$.[10] In other words, '$f$' is a free variable ranging over predicative functions. Indeed one might suggest one of the missing definitions of occurrences of Greek letters as follows (complementing ∗20·02) :

20) $f \in (\phi!\hat{\alpha}) . = . \phi!\hat{x}\,fx \quad Df$

What now of the apparent violation of extensionality at 18)? To begin with we should observe that if one distinguishes classes of classes from classes of functions, and also allows all manner of statements about classes and functions being in classes of either sort, making both $f \in \hat{\alpha}\Phi\alpha$ and $\alpha \in \hat{f}\Phi f$ well formed, no untoward consequences follow. Thus, for example, we need not say that classes of classes, such as $\hat{\alpha}\Phi_1\alpha$ and $\hat{\alpha}\Phi_2\alpha$, have the same classes as members, in this case just the one class Λ, but are

[10]This analysis, and the translation into contemporary terms is also given in Rouilhan's ([10]: 260), but he suggests that it is not supplied in *PM* itself. He does point out the discussion at [*PM*]: 80–81 but then proposes simply ignoring the distinction between functions of classes and functions of functions in the definition ∗20·08 at [10]: 265.

distinct by having "non class" members, namely functions, on which they differ. Classes of classes with the same classes as members cannot simply be "identified", unless one guarantees that anything true of the one class must be true of the other. But having a given function as member is one of those properties "true of" a class, and in that respect the classes will differ. One would indeed have too many classes. We need not accept this unpalatable consequence however, for in fact those two classes of classes will in have the very same functions as members. That is to say, f_1 and f_2 each belong to both of $\hat{\alpha}\Phi_1\alpha$ and $\hat{\alpha}\Phi_2\alpha$. For consider

21) $f_1 \epsilon \hat{\alpha}\Phi_2\alpha$

By $*20{\cdot}08$ this means:

22) $\exists\Phi\ [(\Phi!\alpha \equiv_\alpha \Phi_2\alpha)\ \&\ f_1 \epsilon \Phi!\hat{\alpha}]$

By 20) we have that $f_1 \epsilon \Phi!\hat{\alpha}$ implies $\Phi\hat{x}\, f_1 x$, i.e., $\exists f\ (fx \equiv_x f_1 x\ \&\ \Phi f)$. 22) thus becomes:

23) $\exists\Phi\ [(\Phi!\alpha \equiv_\alpha \Phi_2\alpha)\ \&\ \exists f\ (f!x \equiv_x f_1 x\ \&\ \Phi!f)]$

which is true, taking Φ be Φ_2 and f be f_2. On this account we still have a way of reading sentences that have functions being members of (ϵ) classes of classes $\hat{\alpha}\phi\alpha$, but no awkward results yielding too many classes of classes.

Indeed in general two classes of classes which share all the same classes as members will also have the same functions as members, and so be indiscernible, so the missing theorem of extensionality for classes of classes will obtain:

24) $(\beta \epsilon \hat{\alpha}\Phi\alpha \equiv_\beta \beta \epsilon \hat{\alpha}\Psi\alpha) \equiv (\hat{\alpha}\Phi\alpha = \hat{\alpha}\Psi\alpha)$

To return to our starting point, the proper analysis of an expression $\hat{x}\, fx \epsilon \hat{\alpha}\Phi\alpha$ will arrive at the analysis originally proposed, but making an extra inference in two more steps along the way:

1) $\hat{x}\, fx \epsilon \hat{\alpha}\Phi\alpha$

2) $(\exists g)(\ g!x \equiv_x fx\ \&\ g \epsilon \hat{\alpha}\Phi\alpha)$ using $*20{\cdot}01$

3) $(\exists g)(\ g!x \equiv_x fx\ \&\ (\exists\Psi)(\ \Psi!\alpha \equiv_\alpha \Phi\alpha\ \&\ g \epsilon \Psi!\hat{\alpha}))$ by $*20{\cdot}08$

4) $(\exists g)(\ g!x \equiv_x fx\ \&\ (\exists\Psi)(\ \Psi!\hat{x}h!x \equiv_h \Phi\hat{x}h!x\ \&\ g \epsilon \Psi!\hat{\alpha}))$ by $*20{\cdot}07$

4.1) $(\exists g)(\ g!x \equiv_x fx\ \&\ (\exists\Psi)(\ \Psi!\hat{x}h!x \equiv_h \Phi\hat{x}h!x\ \&\Psi!\hat{x}g!x))$

4.2) $(\exists g)(\ g!x \equiv_x fx\ \&\ (\exists\Psi)(\ \Psi!\hat{x}h!x \equiv_h \Phi\hat{x}h!x\ \&(\exists k)(\ kx \equiv_x gx\ \&\ \Psi!k)))$

5) $(\exists g)(\ g!x \equiv_x fx\ \&\ (\exists\Psi)(\ \Psi!\hat{x}h!x \equiv_h \Phi\hat{x}h!x\ \&\ \Psi!g))$

Sentence 4.1) follows from 4) by the proposed definition based on the remark at page 80 of *PM*. 4.2) then is an application of $*20\cdot01$ to 4.1). Sentence 4.2), which is the proper and full analysis of 1), is logically equivalent to the simpler sentence 5). Now 5) might mistakenly be directly derived from 4) not paying attention to the precise meaning of an expression like $g \in \Psi\hat{\alpha}$, and thinking that one need only apply $*20\cdot02$ to it yielding Ψg. It is certainly true that the class of fs is in the class of classes which are Φ just in case some function Ψ which is coextensive with the function of functions Φ is true of some function g which is coextensive with the function of individuals f, but that derivation goes through the extra step illustrated at line 4.1). Perhaps the fact that the original incorrect derivation still leads to the correct analysis is part of the explanation for why this corner of the system, and the rejected definitions of the Introduction have so far been noticed so rarely.[11] Furthermore, that a theorem of extensionality for classes of classes is derivable, just as one for classes of functions, may be another reason why this whole issue has simply been passed over without notice. Whitehead and Russell do remark that "… classes of classes have all the same formal properties as classes of individuals" ([*PM*]: 189). They did indeed succeed in providing a theory that would allow theorems about classes of individuals, and hence by typical ambiguity, classes of functions, to extend to classes of classes. They succeeded so thoroughly that the complications involved in the distinctions, and in particular, the need for talk of functions being in classes of classes, has just barely been observed, or at least remarked on in print.

In his famous discussion of "Russell's Mathematical Logic", Gödel [4] chastised *PM* for the inadequate introduction of defined expressions, but with just a little charity we may be able to read a great deal of care into the definition of Greek letters α.[12] Those careful definitions require a distinction between classes of functions and classes of classes even though those classes of functions are not needed for the reduction of mathematics that follows. The work of *Principia* is primarily that of constructing an extensional theory of mathematical entities as classes of classes within a background logic of intensional propositional functions.

Acknowledgement. Many people have provided very helpful suggestions and criticisms of this project as it developed from an problem with the number of classes to a counterexample to the theorem of extensionality for classes of classes to its final form as an explication of the obscure definition on page 80 of the Introduction to *PM*. Tony Martin found the puzzle and David Kaplan and Alasdair Urquhart helped me through this progression of responses. Gregory Landini pointed out that he had discussed that passage in the Introduction, and explained his views on this matter to me at the conference from which this volume derives.

[11] Chwistek [3] and Rouilhan [10] are the only two sources in print that I can find that identify this meaning for functions of classes $\Psi\hat{\alpha}$.

[12] Indeed Gödel ([4]: 126) uses the example '$\phi!\hat{u} = \hat{u}[\phi!u]$' as a formula where the order of elimination of the two defined expressions makes a difference to the truth of a sentence. This is almost exactly Carnap's example, but a few years before the publication of [1]. Both were anticipated by Leon Chistwick [4]. See my [8].

References

[1] Carnap, Rudolf: 1947. *Meaning and Necessity.* Chicago: University of Chicago Press. References to second edition, 1956.

[2] Church, Alonzo: 1956. *Introduction to Mathematical Logic*. Princeton: Princeton University Press.

[3] Chwistek, Leon: 1924. The theory of constructive types. (Principles of logic and mathematics.) *Annales de la Société Polonaise de Mathématique. (Rocznik Polskiego Towarazystwa Matematycznego)* II: 9–48.

[4] Gödel, Kurt: 1944. Russell's mathematical logic. In: P. A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, LaSalle: Open Court, 125–153.

[5] Hatcher, William S.: 1968. *Foundations of Mathematics*. Philadelphia: W. B. Saunders.

[6] Landini, Gregory: 1998. *Russell's Hidden Substitutional Theory*. New York and Oxford: Oxford University Press.

[7] Linsky, Leonard: 1983. *Oblique Contexts*. Chicago: University of Chicago Press.

[8] Linsky, Bernard: 2004. Leon Chwistek on the no-classes theory. In *Principia Mathematica. History and Philosophy of Logic* 25: 53–71.

[9] Quine, Willard V.: 1963. *Set Theory and Its Logic*. Cambridge, MA: Harvard University Press.

[10] de Rouilhan, Philippe: 1996. *Russell et le cercle des paradoxes*. Paris: Presses Universitaires de France.

[11] Whitehead, Alfred N. and Bertrand Russell: 1925–1927. [*PM*] *Principia Mathematica*, second edition. Cambridge: Cambridge University Press.

Department of Philosophy
4-115 Humanities Centre
University of Alberta
Edmonton, Alberta
Canada T6G 2E5

E-mail: bernard.linsky@ualberta.ca

A "Constructive" Proper Extension of Ramified Type Theory (The Logic of *Principia Mathematica*, Second Edition, Appendix B)

Allen P. Hazen

Abstract. Russell, in the Introduction and Appendices added to the 1925 second edition of *Principia Mathematica*, suggests a system of Higher-Order Logic which seems to be intermediate in strength between Ramified and Simple Type Theory. It shares a plausible philosophical motivation with RTT, as we show by proving that it has a natural semantic interpretation in which higher-order quantifiers are interpreted "substitutionally," but its criteria of well-formedness are more liberal. Contrary to what has been claimed by Landini, however, it does not validate an unrestricted principle of mathematical induction (as claimed by Russell in Appendix B).

1. History

Russell's metaphysical stance in [46] was an extreme Platonism: you name it and he believed in it, and at least one contemporary reviewer spoke of the book's "scholastic realism." By the time *Principia Mathematica* was written the Platonism was, at the very least, less exuberant. One can speculate as to how much this was a result of the shock of the paradoxes and how much a result of Russell's growing adherence to the traditional British empiricism, but by 1910 Russell was clearly trying to limit his ontological commitments to abstracta, taking as his guiding maxim that one should "[w]henever possible, substitute constructions out of known entities for inferences to unknown entities" [49], leading Gödel [16] to dub Russell's new philosophical stance as "constructivism."

(In itself, that is quite a good name for it, but the logical community now understands the word as referring to programs—very different from Russell's—like those of Brouwer and Bishop. In the two versions of an introductory footnote added when [16] was reprinted in [3] and [39], Gödel noted the difference and characterized Russell's program as "a strictly anti-realistic" or "a strictly nominalistic kind of constructivism." Perhaps—given its influence on and similarity to the programs of Carnap [6] and Goodman [18]—"constructionalism" would be a good term.)

The underlying logic—for purposes of the present discussion we may consider the infamous "Axioms of Reducibility" as non-logical axioms added to it—of the first

edition of *Principia* is the Ramified Theory of Types, described in more detail in section 2. Although set theory is interpreted in it, the abstract entities quantified over by this logic are not sets but rather items of types to which a mind can stand in cognitive relations: *propositions* and concepts, or, in Russell's terminology, *propositional functions*. (This characterization of the ontology of *Principia* is the most serviceable for our purposes, but is probably an oversimplification: cf. the—skippable—last paragraph of this section.) The logic postulates these items only insofar as they can be constructed from a basis of individuals and universals which doesn't have to be specified for mathematical purposes, but which can be assumed, in applications, to be given empirically. As a basis, that is, we assume some number of particulars (items, that is, which could be designated by proper names in an idealized language) and universals—properties and relations—(items which could interpret predicates of appropriate adicities): Russell would claim that, in order for the applied system to be understood, these items should be ones with which we are, in his technical sense, *acquainted*, cf. [47]. Propositions and propositional functions, then, are thought of as structured entities, built up out of the basic particulars and universals in a way which parallels the construction of sentences and open formulas from the names and primitive predicates of a language. The technical mark of the implementation of this conception of propositions and propositional functions in the type structure of the logic is that the formal system of Ramified Type Theory can be given an interpretation under which quantification over other than individuals is interpreted *substitutionally*: details in the next section. It is natural to take the possibility of such a substitutional interpretation of higher-type quantifiers as a criterion of whether a proposed extension of Ramified Type Theory is compatible with Russell's philosophical stance. It seems plausible to interpret Gödel's remark ([16]: 134) that the type-theoretic liberalization of the second edition is "quite unobjectionable even from the constructive standpoint" as signifying that he saw that the liberalized system could be interpreted in this way.

Two comments. First, it does not follow from these considerations that the substitutional interpretation, based as it is on the stock of primitive predicates in a particular formal language, should be thought of as Russell's *intended* interpretation. The substitutional interpretation will validate all *theorems* of the logic, but it is possible to formulate logically consistent existentially quantified sentences which will be false on this interpretation. (Given appropriate choices of vocabulary and individual domain, Axioms of Reducibility will be among the examples.) Such sentences shouldn't be dismissed as logical falsehoods, but should rather be interpreted as contingent assertions implying the existence of universals not expressed by any primitive predicate of our language. (Cf. [20], where it is suggested that this provides a framework for discussing one kind of philosophical question about the completeness of science.)

Second comment: The Axioms of Reducibility are a blatant example of the postulation of unknown entities. Russell could plausibly claim that this was allowed by the escape clause of his maxim. In this instance the substitution of constructions out of known entities is *not* possible, as the Axioms are needed in order to carry out the logicist interpretation of classical mathematics. It is possible to pursue the logicist program

without Reducibility, attempting to derive mathematics within Ramified Type Theory from a simple Axiom of Infinity, and in fact an interesting fragment of arithmetic can be so derived. The fragment, however, though including everyday, gradeschool, arithmetic and weak forms of mathematical induction, is far weaker than (First-Order) Peano Arithmetic, cf. [5] for a summary. Moreover, as remarked by Linsky [30], the entities postulated by Axioms of Reducibility are at least of the same general category as ones we are acquainted with, and, as pointed out by Church [9] and Myhill [33], Ramified Type Theory with Axioms of Reducibility combines mathematical strength with an interesting analysis of the semantic paradoxes.

By the time Russell wrote the Introduction and Appendices to the 1925 second edition of *Principia Mathematica*, his philosophical views had changed again. Unlike Brouwer and Weyl, he had no wish to revise accepted mathematics, and he was happy to republish unchanged the (in modern terms) set theoretic development of it from the first edition: we can see him as ambivalent, still suspicious of but also in part attracted by Ramsey's recommendation in [42] that the complexities of Ramified Type Theory be abandoned in favor of an extensional version of Simple Type Theory. (Extensional Simple Type Theory is a simple and natural version of axiomatic set theory, elegantly compared to Zermelo's system in Quine [41]. The Axioms of Reducibility amount, in effect, to the postulation that the universe of Ramified Type Theory contains a model for Simple Type Theory, and the set-theoretical work of *Principia* is carried out within this model.) He was not, however, completely prepared to abandon the constructivistic program of Ramified Type Theory. In the Introduction and Appendices he wrote for the new edition he sketched a new logic combining some of the constructivistic features of Ramified Type Theory—in particular the fundamental idea that the propositions or propositional functions in the range of a given higher-order quantified variable must be of limited conceptual complexity—with (his interpretation of) the "extensional" interpretation of propositional functions of Wittgenstein and Ramsey.

Russell's conclusion was that this new logic had interesting mathematical strength—he (erroneously) believed that he had derived an unrestricted principle of mathematical induction in it, with an Axiom of Infinity as the only non-logical assumption necessary—but that it was not as strong as simple type theory, and would not suffice for a logicist development of the theory of real numbers. It is tempting to think that the general neglect of the new logic stemmed from this evaluation of its mathematical strength. Russell was not interested in a "revisionary" program, rejecting significant parts of classical mathematics for philosophical reasons: "It might be possible to sacrifice infinite well-ordered series to logical rigour, but the theory of real numbers is an integral part of ordinary mathematics, and can hardly be the object of reasonable doubt" ([50]: xlv). On the other hand, those who *were* interested in such revisionary projects (like Hermann Weyl, whose [56] advocated founding analysis on a basis with some similarities to Ramified Type Theory) were by 1925 committed to other programs, and would not have been interested in something that would have seemed like a minor variant of set theory to them.

In any event, the new logic *was* ignored. None of the contemporary reviews of the second edition that I have been able to examine make any mention of the change to type theory. Indeed, I know of no published evidence that anyone actually *read* Appendix B until Gödel started work on his [16]. (There is evidence that Skolem looked at it … and threw up his hands in horror! Discussing Ramified Type Theory in [53], Skolem discusses general issues in a way that closely follows Russell's introduction, even using one of the same examples, and then attempts to develop a fragment of arithmetic on a predicative basis in a way that owes nothing to Appendix B. It is hard to avoid the thought that Skolem had read the introduction with mounting enthusiasm, had turned to Appendix B, and decided that it would be easier to work things out himself than wrestle with its notational horrors.) Gödel noted the new logic and made brief but clear comments on it, but (perhaps because the comments were lost amidst the wealth of other insights in [16]) no one seems to have noticed his noticing: none of the early comments on Gödel's essay mention it. Myhill [32], though referring to Gödel [16], assumes that Appendix B is based on orthodox Ramified Type Theory.

In 1989 Davoren noticed Gödel's comments and pointed them out to Hazen; they presented a report on their reconstruction of the 1925 logic to a conference in July of 1990, cf. [11]. They claimed that the 1925 logic had a substitutional interpretation, but on the basis of an inadequate proof. (The error was Hazen's; Davoren expressed misgivings before the conference but deferred to her co-author's confidence.) The main contribution of the present paper is to present a correct proof of this claim.

In Appendix B, Russell presented a proof of mathematical induction, using his new logic and an axiom of infinity as the bases for a logicist reconstruction of arithmetic. His proof, however, as Gödel [16] pointed out, is irremediably flawed. (It is also, given the notation, very hard to follow. Gödel gave one possible correction of one error, but the important flaw is separate from this, and hopeless.) Landini [25], however, makes the startling claim that mathematical induction *can* be obtained on this basis; his proof—correct, given his axioms—is a new proof, and not a revision of Russell's attempted proof. Landini's paper is the first rigorous and detailed presentation in print of the 1925 logic. He claims that it has a substitutional interpretation (he has confirmed in correspondence that this is what he means when he refers to the system's "nominalistic semantics"), but without proof. With one major difference, his formalization of the logic is equivalent to that given by Davoren and Hazen (and restated in section 3 below): Landini adds an axiom scheme of extensionality. The second main point of the present paper is to argue (section 4) that this axiom scheme, which is required for his proof of mathematical induction, is incorrect on the substitutional semantics, and so unacceptable from Russell's "constructivistic" point of view. (There is also the minor difference that Landini does not include abstracts in his formal language.)

SKIPPABLE LAST PARAGRAPH: The conception of propositions and propositional functions as structured entities suggested above (and discussed in greater detail in [22] is reasonable from a modern standpoint, and suffices for the motivation of the Ramified Type Theoretic logics of (first and second edition) *Principia*. It would have seemed less reasonable to Russell. We modern logicians are happy with structured entities

(in, e.g., syntax) in part because we know that the nuts and bolts for putting them together—pairing, formation of finite sequences—is available at the safest and most elementary levels of set theory. Russell, who was trying to reconstruct set theory on the basis of a theory of propositional functions, was not in a position to presuppose this material. (And, more generally, even such simple constructs as ordered pairs would have presented foundational issues to a set theorist at the beginning of the 20th Century: Peirce's "dyad … a mental diagram consisting of two images of two objects …" had not yet been reduced by Wiener and Kuratowski! Cf. [40]: section 53.) Russell (at least in part, one assumes, because of this problem) tried, in the period in which the first edition of *Principia* was completed, to avoid ontological commitment to propositions as structured entities altogether. This was the (a?) point of his "multiple relation" theory of judgment: instead of construing judgment (belief), etc., as a binary relation between a judging mind and a structured proposition, he construed it as a many-place ("multiple") relation between the mind and the various simple constituents of the proposition. (Strictly speaking: between the mind and the various entities, particular and universal, which would have been constituents of the proposition if there had been propositions!) He advocates the multiple relation theory in the introductory chapter on the theory of types in (the first edition of) *Principia*, writing

> Owing to the plurality of the objects of a single judgment, it follows that what we call a "proposition" (in the sense in which this is distinguished from the phrase expressing it) is not a single entity at all. That is to say, the phrase which expresses a proposition is what we call an "incomplete" symbol; it does not have a meaning in itself, but requires some supplementation in order to acquire a complete meaning. (p. 44)

Now, "incomplete symbol" is Whitehead and Russell's term for things like definite descriptions and class abstracts: pieces of notation to be explained by contextual definitions (and, indeed, the passage just quoted has a footnote referring the reader to the chapter on such notations). The obvious implication is that quantification over propositions (and so, also, propositional functions) is to be explained away as a *façon de parler*, much as one might explain quantification over rationals (and atomic formulas containing variables for rationals) as shorthand for pairs of quantifiers over integers (and expressions containing double the number of variables). But, as Church [9] complains,

> … the contextual definition or definitions that are implicitly promised by the "incomplete symbol" characterization are never fully supplied, and it is in particular not clear how they would explain away the use of bound propositional and function variables. If some of the things that are said by Russell in IV and V of *Introduction to the second edition* may be taken as an indication of what it intended, it is probable that the contextual definitions would not stand scrutiny. (footnote 4; when the paper was reprinted Church replaced this footnote with a more cautiously worded one.)

In fact there is a serious technical problem in giving appropriate contextual definitions. If all the propositions referred to or quantified over were of the same form—e.g., subject/predicate, composed of one monadic universal and one particular—the contextual definitions would be easy to state. Atomic formulas of the form $J[ap]$ ("a judges that p") with a binary judgment relation could be replaced by atoms of the form $J[abF]$ ("a bears the ternary judgment relation to the subject, b, and predicate, F, of p"), and quantifiers over propositions, $\forall p$ or $\exists p$, could be replaced by pairs of quantifiers, $\forall x \forall F$ or $\exists x \exists F$. The underlying logic would be ordinary First-Order Logic (perhaps double-sorted). For the theory of propositions (and, more importantly, propositional functions) to serve as a foundation for mathematics, however, quantifications could not be restricted to propositions of a single form, or even of a limited number of constituents. The "multiple relation" of judgment would have to be, not merely multiple, but *multigrade*, and, what is technically far harder to handle, the quantifiers of the underlying logic would have to be able to bind "vectors," of varying length, of variables. Such a logic is described (and its application to the multiple relation theory of judgment briefly discussed), in [54], where it is proven to be unaxiomatizable (that is, that its set of valid formulas is not recursively enumerable). I would suggest, therefore, that Russell was serious about the nonexistence of propositions (and propositional functions), and thought that in principle (well-known philosophical weasel words, those!) it should be possible to think of them as "incomplete symbols" (this is also argued in [26]), but for good technical reasons was unable to carry out their reduction in detail. That said, however, the conception of propositions (etc.) as structured entities seems like a clearer way of motivating the system he actually published.

2. Ramified Type Theory and Its Semantics

In order to facilitate comparison with the 1925 logic, this section presents a version of Ramified Type Theory ("IRT"—Edition *I*, *R*amified *T*ypes), generally following Church [9] but formulated with abstracts (and rules of β-conversion) instead of comprehension axioms. The second subsection discusses the status of extensionality principles and the motivation for using abstracts, and the third presents the "substitutional" interpretation. A fourth takes an excursion into proof theory.

THE LOGIC IRT: Since Ramified Type Theory adds formal complications to Simple Type Theory, it is expositorily convenient to start with the latter. We assume entities to be organized into a system of *types* of the following sort. There is a type, i, of urelements or *individuals*, and for every finite sequence of types $t_1, \ldots, t_n$ a type $(t_1, \ldots, t_n)$ of *propositional functions* taking a sequence of arguments of the given types $t_1, \ldots, t_n$: n-adic *relations* (*properties* in the case $n = 1$) or, if we assume extensionality, n-adic *relations-in-extension* (*classes* in the case $n = 1$). We include also the case of zero-length sequences, yielding a type () of *propositions* or, assuming extensionality, truth-values.

A conventional formal language embodying such a conception would be based on a many-sorted quantificational language, with all variables and constants having assigned types. There would be alphabets of free and bound variables for each type, interpreted as ranging over entities of that type, and, optionally and depending on application, constants of some types, interpreted as denoting specific entities of those types. Free variables (and constants) of type () standing alone count as formulas; all other atomic formulas are of the form $a[b_1, \ldots, b_n]$, where $b_1, \ldots, b_n$ are free variables (or constants) of the respective types $t_1, \ldots, t_n$, and a is a free variable (or constant) of type $(t_1, \ldots, t_n)$. Non-atomic formulas are formed in the usual ways, by combining formulas with connectives or by going through a formula, replacing every occurrence in it of some one free variable with occurrences of a bound variable (not already occurring in the formula) of the same type, and prefixing a quantifier with this bound variable as operator variable. (In other words: I adopt Hilbert and Gentzen's convention, under which bound and free variables are notationally distinct, bound variables do not occur free in well-formed formulas, and quantifiers with the same operator variable never have overlapping scopes.) *Sentences*, as usual, are formulas not containing free variables, *quasi-formulas* things like formulas except for containing unbound occurrences of bound variables in places where they ought to have free variables.

There is an optional added feature for this language which is convenient for some purposes and trivial on some (but not all) assumptions. A formula containing free variables is naturally thought of as expressing a propositional function. In order to ensure that the language contains a term for each propositional function expressed by one of its formulas, we may add *abstracts*. Where A is a quasi-formula containing free occurrences of no bound variables other than $x_1, \ldots, x_n$, of types $t_1, \ldots, t_n$ respectively, $\{x_1, \ldots, x_n : A\}$ is an abstract of type $(t_1, \ldots, t_n)$, and may occur in any environment in which a free variable or constant of that type may occur (and occurrences of $x_1, \ldots, x_n$ in an occurrence of $\{x_1, \ldots, x_n : A\}$ in a formula are counted as bound in the formula). We can make our machinery uniform by saying that, where A is a formula, $\{A\}$ is an abstract of type (). As is well-known (cf. [41] for discussion), abstracts are eliminable by contextual definition if extensionality is assumed. They are not in general eliminable without this assumption, and in some applications give a useful increase in expressive power. Note that every occurrence of a bound variable in an abstract is either bound in the formula following the colon or is an occurrence of one of the variables preceding the colon. Repeated formation of abstracts, or quantification of formulas containing abstracts, can produce occurrences of *quasi-abstracts*, expressions like abstracts but containing free variables which are neither operator variables of the abstract nor bound by operators within the abstract.

The formal language is rich enough that, on its familiar set-theoretic interpretation with infinitely many entities of type i, completeness is not to be hoped for. A basic axiomatization should, however, include at least standard quantificational logic for each type, and some embodiment of the motivating idea that formulas express propositional functions and sentences propositions. So: start by postulating the usual

natural deduction or sequent-calculus rules for the connectives and quantifiers (the quantifier rules for variables and constants of each type), as in [14]. These rules by themselves, however, do not allow us to prove that there is a propositional function expressed by any given formula. As stated so far, the quantifier rules allow free variables and constants (if any) to be substituted for the bound variables of higher types: what is needed is some way of letting *formulas* be substituted.

There are two general approaches possible here. One is to postulate axioms of *comprehension*. This is satisfactory if extensionality is assumed, but seems inadequate without this assumption: a propositional function holding of all and only those (sequences of) items which satisfy a formula, which is what an axiom of comprehension gives, may not be the propositional function expressed by that formula. In the non-extensional case it is possible to get around this difficulty by enriching the language with a non-truth functional connective of *propositional identity* and formulating the comprehension axioms with this connective, rather than material equivalence, as in [8] and [10] and (following Church) [1] and [2]. A simpler alternative, which will be followed here, is to supplement the statement of the quantifier rules, for types other than i, with a definition of what it means to substitute a formula for a variable, as in (the first part of) [8]. Given the complexity of this notion, it is more convenient to introduce abstracts as auxiliary notation.

Postulate, then, rules of β-conversion to handle abstracts: formulas obtained from one another by β-conversion (sequents obtained from one another by replacing a formula by one obtained from it by β-conversion) are equivalent, and may be inferred from one another. (This is an effective rule: since we are working in a typed system, and indeed with one having only a small fraction of the hierarchy of types available in systems of typed λ-calculus, β-convertibility is decidable. At considerable cost—derivations get much longer—we could get the same effect by postulating simple rules of *abstraction* and *concretion*.) Then, in stating the quantifier rules, allow bound variables to be substituted for abstracts (though not, of course, quasi-abstracts), as well as free variables and constants, of their type.

[ASIDE: Even if the hypotheses and conclusion of a derivation (to put it in natural deduction terms: in sequent-calculus terms, even if the formulas in the endsequent of a proof) are free of abstracts, a derivation using these rules can contain formulas essentially containing abstracts: that is, formulas which contain abstracts even when reduced to β-normal form. The irreducible abstracts in these formulas will occur as terms of atomic quasi-subformulas with variables as predicates. Consider, for example, the following derivation of $\exists X \exists \Phi(\Phi[X] \ \& \ X[a])$ (with variables of types Φ : ((i)), X : (i), and a : i) from $F[a]$:

1.	$F[a]$	Hypothesis
2.	$(F[a] \ \& \ F[a])$	1, & introduction
3.	$(\{x : F[x]\}[a] \ \& \ F[a])$	2, β-abstraction

4.	$(\{X : X[a]\}[\{x : F[x]\}] \;\&\; F[a])$	3, β-abstraction
5.	$\exists\Phi(\Phi[\{x : F[x]\}] \;\&\; F[a])$	4, $\exists$ introduction
6.	$\exists\Phi(\Phi[\{x : F[x]\}] \;\&\; \{x : F[x]\}[a])$	5, β-abstraction
7.	$\exists X\exists\Phi(\Phi[X] \;\&\; X[a])$	6, $\exists$ introduction.

The abstracts introduced in 3, 4 and 6 are part of the machinery: those introduced in 3 and 6 would not occur if the $\exists$ introduction rule had been formulated in terms of a rule of substitution of formulas for variables instead of taking a detour through abstracts. Formula 5, however, is in β-normal form and contains an irreducible abstract. 7 can, of course, be derived from 1 in other ways, but this seems the least clumsy. The moral seems to be that one *can* treat abstracts as mere auxiliary notation to simplify the statement of the rules of deduction, but that it is more convenient to grant them full citizenship, extending the expressive resources of the language to allow sentences like 5. END OF ASIDE.]

Extensional Simple Type Theory is a reasonable formulation of set theory: not as strong as Zermelo's, but working mathematicians' systematic use of capital and lower case letters, and distinguishing locutions like "family of sets," suggest much mathematical thinking is in practice type-theoretic. The non-extensional variant, however, is inadequate as a theory of propositions and propositional functions: it succumbs to the semantic and epistemic paradoxes. Russell moved to a theory making finer discriminations of type. Church [9] refers to the types of Ramified Type Theory as *r-types*.

Following the account in [9], we once again start with a type i for individuals. As in the Simple Type Theory, entities of other types may be distinguished according to the types of arguments they take, but now further distinctions must be made, reflecting the differing structural complexities of the propositional functions (or propositions) falling within a single (Simple) type. Start with those entities taking only individuals as objects (including those abstract entities taking no arguments, the propositions). Each type subdivides into a hierarchy of *ramified types* (*r-types*, for short) according to the complexity of their structure or constituency. Thus, instead of simply having types of properties, dyadic relations, triadic relations, and so on, (i), (i, i), $(\mathsf{i}, \mathsf{i}, \mathsf{i})$, and of propositions, $()$, we have, corresponding to each such simple type a series of types, $(\mathsf{i})/1$, $(\mathsf{i})/2$, $(\mathsf{i})/3$, …, $(\mathsf{i}, \mathsf{i})/1$, $(\mathsf{i}, \mathsf{i})/2$, $(\mathsf{i}, \mathsf{i})/3$, … and so on. It is convenient, though not faithful to the letter of Russell's exposition, to interpret each of these hierarchies *cumulatively*, so that, for example, an entity of type $(\mathsf{i}, \mathsf{i})/3$ is counted as also being of types $(\mathsf{i}, \mathsf{i})/4$, $(\mathsf{i}, \mathsf{i})/5$, and so on. With the types so far listed, we have *Ramified Second-Order Logic*, as described in sections 58 and 59 of Church [7]. The numeral after the slash in a type designation is said to give the *level* of the type, or of the entities of that type.

To get *ramified Third-Order Logic*, we add types for entities taking arguments of at least one of the non-individual types recognized in ramified Second-Order Logic, where, once again, the entities taking (a sequence of) arguments of a given (sequence of) types are subdivided into levels according to their structure or constituency. Thus we will have types $((\mathsf{i})/1)/1$, $((\mathsf{i})/1)/2$,…, of properties taking as arguments level-

one properties of individuals, types (i, (i)/3)/1, (i, (i)/3)/2,..., of relations between individuals and level-three properties of individuals, and so on. By the cumulativity of levels for Second-Order types, an (i, (i)/3)/1 can relate an individual to an (i)/1 or an (i)/2; levels of Third-Order types are also cumulative, so an (i, (i)/3)/1 will be counted to be an (i, (i)/3)/2 as well. In Third-Order Logic (and at higher types) it is useful to define a further notion, the *badness* (Church [9] uses Russell's term *order*, but this conflicts with the modern terminology of "n-th-Order Logic") of a type. Individuals are defined to have badness 0; for other types the badness is the sum of the "worst" allowable argument for an entity of the type with the level of the type. Thus the type (i, (i)/3)/1 has badness 4 (= 3, the badness of those (i)/3 entities which are not (i)/2's, +1, the level of the relational type), and (i, (i)/3)/2 has badness 5.

Ramified Fourth-Order Logic adds in addition types of entities taking as arguments entities of at least one Third-Order type, and the types of a full system Ramified Type Theory, such as our IRT, are all those belonging to ramified n-th-Order Logic for some natural number n.

The *vocabulary* of IRT will contain (free and bound) variables of all these types, and perhaps constants of some; constants and variables will be said to have the *badness* of their type. (*Principia Mathematica*, which was concerned with purely logical and mathematical issues, includes no non-logical constants in its formal language. Its authors regarded this language as a sort of skeletal language, one which would be able to express the contents of arbitrary thoughts once it was supplemented with appropriate non-logical constants standing for items of direct acquaintance.) *Atomic formulas* are of the form $a[b_1, \ldots, b_n]$, where the b_i are terms (constants, variables, or abstracts) of types suitable as arguments for items of a's type: note that the badness of the predicate of an atomic formula will always be greater than the badness of any of its arguments. The set of *well-formed formulas* is specified by the usual sort of recursion. Its *deductive apparatus* will include, as before, the usual rules of quantificational logic for all types, and rules of β-conversion. The level-aspect of r-types is supposed to reflect the constructional complexity of propositions and propositional functions, and this is what motivates the logic's specification of what abstracts may be substituted for the variables of a given (non-individual) type. One preliminary technical definition will be useful before stating this specification. We may state the:

> FUNDAMENTAL SPECIFICATION: In the application of the Quantification and β-conversion rules, an abstract $\{x_1, \ldots, x_n : A\}$ is substitutable for a variable of type $(t_1, \ldots, t_n)/j$ and badness k (where $k > j$) if and only if
>
> 1. the abstraction variables $x_1, \ldots, x_n$ are, in order, of types $t_1, \ldots, t_n$,
> 2. no free variable or constant occurring in A is of badness $> k$, and
> 3. every bound variable occurring in A is of badness strictly $< k$.

ABSTRACTS AND EXTENSIONALITY: Many applications of IRT (including Russell's intended application for Ramified Type Theory) are incompatible with assuming ex-

tensionality, but an extensional variant might be more convenient for mathematical applications. *Imposing* extensionality axiomatically is slightly more complicated than it is in Simple Type Theory and other standard set theories. In Simple Type Theory one can postulate, for each type t, a single axiom of extensionality (with variables of types $\Phi : ((t))$, $X : (t)$, $Y : (t)$ and $x : t$):

$$\forall X \forall Y (\forall x (X[x] \leftrightarrow Y[x]) \rightarrow \forall \Phi (\Phi[X] \leftrightarrow \Phi[Y]))$$

In IRT this is not possible, the problem lying with the quantifiers of the two higher types. There are infinitely many types, $(t)/1$, $(t)/2$, etc., of propositional functions taking items of type t as arguments, so infinitely many axioms with the $\forall X \forall Y$ quantifiers rewritten in these different types will be needed. Worse, each of *these* axioms will have to be reduplicated infinitely many times, with the $\forall \Phi$ rewritten in the different types capable of taking arguments of the type of the $\forall X \forall Y$ quantifiers. (In other words, the axiom of extensionality, already schematic in a type-theoretic context, becomes triply so!)

So much for extensional identity at higher types. In Simple Type Theory one can also define identity of individuals by the formula

$$\forall X (X[a] \leftrightarrow X[b])$$

As Russell pointed out, however, this will not suffice in a Ramified logic: whatever level is chosen for the $\forall X$ quantifier, the corresponding formula with a higher-level X will not be derivable from it (this non-derivability is rigorously established in Myhill [32]). So yet another infinite set of extensionality axioms will have to be added:

$$\forall X (X[a] \leftrightarrow X[b]) \rightarrow \forall Y (Y[a] \leftrightarrow Y[b])$$

(with variables of types $X : (\mathrm{i})/1$, $Y : (\mathrm{i})/n$ for $n > 1$, and $x : \mathrm{i}$). The resulting extensional system is perhaps the best version of Ramified Type Theory to use in investigating the mathematical power of Ramified Types without the Axiom of Reducibility. It is interpretable in our basic IRT by making appropriate restrictions on all higher-order quantifiers.

Without extensionality, the use of comprehension axioms will not give as flexible a system as the use of abstracts. Church's [9] is something of a *tour de force*, demonstrating that it is possible to use Ramified Type Theory to give an interesting analysis of (at least some) semantic paradoxes while using a variant of the logic with only comprehension axioms, but for other purposes the expressive power of abstracts is desirable. For example, the use of abstracts allows, what comprehension axioms do not, a natural identification of *the* propositional function *expressed* by a given formula of the formal system.

SEMANTICS WITH SUBSTITUTIONAL QUANTIFIERS: Given that the motivating idea of propositions and propositional functions has them structured in a way that parallels the construction of the formulas of the formal language, the idea of interpreting the higher-order quantifications of Ramified Type Theory substitutionally was a natural one. An interpretation of this kind was given by Fitch [13], and substitutional inter-

pretations have been discussed by a number of writers since. The clearest discussions are in Parsons' [34, 35, 36, 37].

There are both conceptual and technical difficulties with such a semantics, which we can either ignore or finesse. One conceptual problem, leading to doubts that a purely substitutional semantics should be thought of as the correct or intended interpretation, has been mentioned in the first section of this paper: there can be no guarantee that (empirical) discoveries will not lead us to supplement our language with new predicates expressing universals that are at present unknown. It is also not clear how we should formulate the semantics of higher-order predicate constants, such as those expressing intensional notions like *meaning* or *belief*. This problem can be ignored in the context of a comparison with Russell's 1925 logic, for such constants were explicitly excluded from the later logic. Finally, there is the problem of referentially interpreted quantification over *individuals*: if there are nameless individuals in the domain interpreting variables of type $\mathfrak{i}$, it will in general not be possible to interpret quantifications of other types in terms of *closed* substitution instances. This last technicality was pointed out by Parsons; for simplicity we may limit consideration to interpretations of IRT over models with domains of *named* individuals.

A *substitutional model* for IRT (with no predicate constants of types other than $(\mathfrak{i}, \ldots, \mathfrak{i})/1$), then, is simply a structure in the sense of the model theory of First-Order Logic, interpreting the predicate constants, in which all individuals are named. As one would expect, an atomic *sentence* (the semantics deals only with closed formulas!) is true if and only if the tuple of objects named by its terms are in the extension of its predicate. It will slightly simplify the exposition (and, since the existence of β-normal forms is decidable, etc., it is harmless) to assume that all sentences are in β-normal form: abstracts occur only as arguments to bound predicate variables.

Then truth of sentences in general can be defined by the usual sort of recursion:

- a negation is true iff the negated sentence is not true;
- a conjunction (disjunction) is true iff both of its conjuncts (one or more of its disjuncts) is true;
- similarly for any other connectives you want to include;
- a universal quantification is true iff all of its *instances* are true; and
- an existential quantification is true iff at least one of its instances is true, where
- an instance of a quantification with quantified variable of type $\mathfrak{i}$ is any sentence formed from it by deleting the initial quantifier and replacing all occurrences of the variable bound to it with occurrences of some one name, and
- an instance of any other quantification is any sentence formed from it by deleting the initial quantifier and replacing all occurrences of the variable bound to it with occurrences of some one *closed* abstract and (if necessary) normalizing.

Intuitively it is clear that this semantics explains the truth value of every *complex* sentence by reducing it to a question of the truth values of *simpler* ones. This notion of reduction is often dramatized with the metaphor of a semantic *game*. To determine the truth value of a sentence, we give it to two omniscient players, a *Proponent*, who will win if the game ends with a true atomic sentence, and an *Opponent*, who will win if the game ends with a false atomic sentence. At each stage of the game, if the sentence in play is a conjunction or universal quantification, the Opponent gets to choose a conjunct or instance, trying for a false one. If the sentence in play is a disjunction or existential quantification, the Proponent gets to choose, trying to choose a true disjunct or instance. If it is a negation, they switch sides and play with the negated sentence; rules for other connectives are readily specified. The starting sentence is true if the Proponent can win, false if the Opponent can.

In technical terms, the well-definedness of the semantics depends on the fact that the relevant reducibility relation is well-founded. In terms of the game-metaphor, the requirement is that the game will always *terminate*, whatever sentence the players start with: no sequence of moves, each move reducing the sentence in play to one of those its truth value is defined to be dependent on, can go on forever. This is perhaps obvious enough without proof, but the proof is, in any event, quick and easy. We define an ordinal measure of complexity, or *grade*, on sentences and their quasi-sub-formulas as follows:

- the grade of an atomic sentence or quasi-formula with a predicate constant as its predicate is 0
- the grade of an atomic quasi-formula with a predicate variable as its predicate is $\omega \cdot k$, where k is the badness of the variable
- the grade of a negation is the grade of the negated sentence or quasi-formula $+ 1$
- the grade of a conjunction (etc.) is 1 more than the grade of whichever conjunct $(\ldots)$ has the higher grade
- the grade of a quantification is the grade of the matrix $+ 1$.

It is easy to see that the sentences to which the semantics "reduces" the truth value of a given non-atomic sentence will always be of lower grade. The only difficult case is that of a quantification with a quantified variable of non-individual type. Since the arguments to a predicate variable must have lower badness than the variable itself, we may ignore occurrences (of the quantified variable to be replaced) as arguments in atomic quasi-subformulas: their replacement with abstracts will not immediately affect the grade of the sentence, and when they do come into play (at a later stage of the reduction, after the predicate has been substituted for) they will not cause any damage. There remains the case of a predicate variable occurring as a predicate. Since, however, it is to be replaced with a *closed* abstract, every variable occurring in

the abstract, as well as variables standing as (or occurring within abstracts standing as) arguments to it must be of strictly lower badness. The (quasi-) formula formed when the abstract is β-reduced will, therefore, have a lower grade than the original atomic quasi-formula, and so (by induction on the construction of the sentence) the whole sentence will have its grade reduced.

Note that the "substitutional" semantics defined for IRT is an "honest" one, unlike that claimed for Simple Type Theory in [28]. As stressed earlier, when we start with a particular set of primitive predicates, there will in general be existentially quantified sentences that are consistent (with the deductive system for IRT) but false on the substitutional interpretation as defined: such sentences may be thought of as assertions that the set of predicates of the language is inadequate. On the other hand, every theorem of IRT will be true on the semantics: the soundness of the deductive apparatus can be established by the usual sorts of argument. (Completeness, of course, is another matter: as is well-known—cf. [12]—completeness on a substitutional semantics is not to be hoped for without infinitary inference rules!) It is, in other words, *consistent* with the acceptance of the deductive apparatus of IRT to believe that one's language has enough predicates to express all the ways in which individuals resemble and differ from one another. Leblanc's substitutional interpretations, in contrast, are related to non-standard models (in the sense of Henkin [24]) of Simple Type Theory, and may require the addition of infinitely many new primitive predicates to a given language in order to verify the theorems of the system. Leblanc, in other words, obtains substitutional semantics only by *postulating* an unknown language extending the one he starts with: surely a case of theft rather than honest toil! On the other hand, the semantics of Leblanc & Weaver [29] (and also that presented in Grover [19] in the same volume) is close to that presented here.

A NOTE ON CUT-ELIMINATION: The complexity measure used here is closely related to Gentzen's notion of *grade* in [14] , which was employed in proving his *Hauptsatz* for First- Order Logic, and it is almost true that, after extending our definition to cover formulas containing free variables, we can use it to prove a cut-elimination theorem for IRT. There is a slight complication, however: since the abstracts substitutable for a bound predicate variable in applications of the quantifier rules are allowed to contain free variables of the same badness, the simple definition of grade used above for semantic purposes allows open instances to exceed a quantification in grade. For example (where P and Q are purely atomic sentences of grade 0, a is an individual constant or variable, and the predicate variables X and R are of types (i)/1 and (i, i)/1 respectively),

$$(P \ \& \ (Q \ \& \ \forall x Rxa))$$

is of grade $\omega + 3$, but, after abstracting to

$$\{y : (P \ \& \ (Q \ \& \ \exists R \ \forall x Rxa))\}[a],$$

we may infer

$$\exists X(X[a]),$$

which is only of grade $\omega+1$, from it. For proof-theoretic purposes, we need a slightly more refined notion of grade. We may once again allow atomic sentences (and open formulas as well, this time) with constants as predicates to have grade 0, and truth functional connectives and quantifiers over *individual* variables may increase grade by 1 as before. The remaining two clauses of the definition are changed to

- the grade of an atomic (quasi-)formula with a variable as predicate is $\omega\cdot(2k-1)$, where k is the badness of the predicate variable, and
- the grade of a quantification with a non-individual variable in the initial quantifier is $\omega\cdot 2k$, where k is the badness of the variable, if the grade of the matrix is less than $\omega\cdot 2k$, and otherwise is 1 greater than the grade of the matrix.

With this altered definition, the limit of the grades of well-formed formulas of IRT remains ω^2, but now open as well as closed instances of quantifications will be of lower grade than the quantifications themselves. (PROOF: *Case 0*: the quantifier removed has a variable of other than the highest badness in the formula. Then the quantifications of the variables of higher badness will determine the grade of the matrix, and removal of the initial quantifier will reduce the grade of the whole by 1. *Case* 1: the quantifier removed has a variable of higher badness than any bound variable in the matrix, so the quantification has a grade of the form $\omega\cdot 2n$. Since bound variables in the abstract replacing the variable will also be of lower badness, the maximum grade for the instance will be $\omega\cdot(2n-1)+$ some natural number. *Case* 2: there are additional bound variables of the same badness in the matrix. Note that, on the assumption that the formula is in β-normal form, no occurrence of the abstraction operator can bind a variable of the highest badness in the formula. If all the occurrences of the variable to be replaced are inside the scope of quantifiers binding some of these additional variables, then the grade of the matrix will not be affected by the replacement, and the grade of the whole will be reduced by 1 when the initial quantifier is removed. If, on the other hand, there are occurrences of the variable to be replaced that are not in the scope of such quantifiers, the largest subformulas of the original matrix containing such occurrences but not containing any quantifiers with variables of the same badness will, after the replacement, come to have a grade no higher than $\omega\cdot(2n-1)+$ some natural number, and so, from the point of view of calculating the grade of the whole instance, will be dominated by the subformulas, truth functionally connected to them, that *do* contain quantified variables of highest badness. Thus in this case, as in case 0, the grade of the instance will be 1 less than that of the quantification. END OF PROOF.)

With this revised definition of grade, then, a *Hauptsatz* for IRT can be proven by reconstruing Gentzen's induction on grade as transfinite induction, and otherwise taking over his proof unchanged. (The resulting *Hauptsatz* is for a slightly abstract

formulation of the logic, *identifying* β-equivalent formulas and ignoring steps of β-abstraction and reduction.) The provability of cut-elimination for Ramified Type Theory has been known at least since Lorenzen [31], but I know of no readily accessible treatment in English. Discussion of Ramified Type Theory (and of Ackermann's type-free system) in Schütte [52] was dropped from the English translation, but there is a pseudo-translation (replacing Schütte's idiosyncratic variant of sequent-calculi with Beth–Smullyan tableaux) in Toledo [55].

3. The 1925 Logic

In this section we describe the 1925 logic (BMT for "Appendix *B*, *M*odified *T*ype Theory). The first subsection describes the motivating ideas and the divergences from IRT. The second (written to be independently readable) gives a formal description of BMT. BMT as presented in the second subsection, though motivated by an extensional conception of propositional function, contains no extensionality principles: the third subsection discusses the degree to which extensionality can be expressed by added axioms or rules. The fourth subsection proves the existence of a substitutional semantics.

THE INFLUENCE OF RAMSEY AND WITTGENSTEIN: Russell's earlier logical work had been directed toward the construction of a general intensional logic, one in which such notions as the *cognitive* relations between a mind and the propositions it understands and believes could be expressed. The type structure of IRT, according to which a given higher-order propositional function could apply only to arguments of limited badness (and the badness of a predicate is always greater than that of the arguments to which it is applied) is natural and well-motivated: it allows the type theory by itself, without *additional* apparatus, to defuse the semantic and *intentional* (epistemic) paradoxes as well as the set-theoretic ones. Consider, for example Buridan's scenario, in which Socrates (by hypothesis an infallible and instantaneous logician) reads with understanding the following graffito:

(*) Socrates does not believe this proposition

(and let us interpret "proposition" in a Russellian rather than Buridanian way). Socrates, ever questioning, considers the proposition, and sees that if he were to decide it was true and start to believe it, this would render it false: which, since he recognizes it, means he cannot rationally believe it. On the other had, unless and until he comes to believe it, it is true, and this too he recognizes. Assuming, as we have, that his beliefs are formed with perfect rationality, he must both believe it and not believe it.

Most modern treatments of this take the "this proposition" to be a definite description, short for something like "the proposition expressed (in English) by this inscription." This allows at least two distinct resolutions:

(a) we may (following Russell's discussion of, e.g., the Epimenides) point out that when the definite description is given its Russellian contextual definition, a propositional quantification occurs, and Ramified Type Theory yields the result that the proposition is non-paradoxically false—whatever level (badness) we assign to the quantification in it, the proposition itself will be of the next level up, so, in the sense of the quantifier, there will be no proposition expressed by the inscription, or

(b) we may instead choose to "ramsify, not ramify," and focus attention on the systematic ambiguity of "expressed," appealing to the doctrine of "levels of language" to conclude that no proposition at all is expressed by the inscription, for the relation of *expression* holding between sentences of any language and propositions can only be expressed by a predicate of a higher language.

Neither of these approaches require that propositional functions be limited in the *levels* (as opposed to the simple types) of the arguments they take. It is, however, far from obvious that "this proposition" should be analyzed as a definite description or as involving a reference to the inscription. "This," after all, is Russell's favorite example of a logically perfect name, and if Socrates, having understood the inscription (and being acquainted with each of the constituents of the proposition expressed), is actively considering the proposition, then this is as good a candidate as one could hope for an example of direct acquiantance with a proposition. Suppose, therefore, that we interpret the "this proposition" as directly referential. There is then no mention of the expression relation, and no quantification over propositions. The only (obvious) remaining way to defuse the paradox is to postulate a type-ambiguity in *believe*. Since there are many types of propositions, the type structure of IRT forces us to suppose that there is a corresponding hierarchy of belief relations, and the paradox is dispelled by noting that no belief relation can apply to a proposition of which it is a constituent.

In a certain sense (Russell, in the Introduction to the second edition of *Principia*, uses these very words), however, low-level propositional functions can apply to high-level arguments. Suppose we want to say of a $(\mathsf{i}, \mathsf{i})/1$ relation, R, expressed by a predicate constant, that it is, say, symmetric. We may do so by asserting a purely First-Order sentence:

$$\forall x \forall y (Rxy \rightarrow Ryx).$$

Replace the predicate by a predicate variable and we get a formula defining a $((\mathsf{i}, \mathsf{i})/1)/1$ property of relations. But now suppose we wanted to say of some high level binary relation of individuals—perhaps an $(\mathsf{i}, \mathsf{i})/79$—that *it* was symmetric. Whatever horrendous formula we would have to write to express the relation, the additional material needed to say it was symmetric would be precisely the same as the material used in saying that R was symmetric: the conditional operator and a couple of quantifiers with variables ranging over individuals. *Officially*, we have a hierarchy of symmetry properties: the $((\mathsf{i}, \mathsf{i})/1)/1$ first defined, the $((\mathsf{i}, \mathsf{i})/79)/1$ we (in effect) defined when we said the $(\mathsf{i}, \mathsf{i})/79$ was symmetric, and so on. (And, though the cumulative interpretation of levels allows us to drop some of these properties in favor of others, the fact that there is no highest level of binary relations of individuals means that no *one*

symmetry property can be ascribed meaningfully to all such relations.) Still, although the theory of types implicit in IRT insists that these are different, there is a clear sense in which they have much in common, in which they are defined in the same way. If we didn't have some *other* reason for distinguishing between properties of relations taking different levels of relation as arguments, it would be tempting to identify them . . .

By about 1920 Russell, under the combined influence of Wittgenstein and of such psychological behaviorists as Watson, had largely abandoned the notion of belief, etc., as relations between minds and propositions. Appendix C, added to the second edition of *Principia*, sketches a defense of the new view, on which there are no intensional primitives, and the only higher-order properties (relations) recognized are those definable by logical means from (ultimately) the First-Order primitives. (Readers who recall the jacket blurb, "This is the book that has meant the most to me," from W. V. Quine on the back cover of Whitehead & Russell [58] will see in Russell's ontological development a precedent for Quine's later "flight from intension.") The net effect of the psychological speculation is to suggest a reduction of intentional notions to semantic ones, somewhat in the manner of more recent speculations in cognitive science about "the language of thought." As for the semantic paradoxes, they can be defused by means of the doctrine of "levels of language" (as Gödel [16] hints *via* references to Ramsey and Tarski). It seems fair to attribute this line of thought to Russell, as he had already given a very clear statement of the levels of language idea at the end of his preface to Wittgenstein's *Tractatus*.

One initially obscure-sounding statement of the new view that Russell gives in the Introduction to the second edition, "Propositional functions occur only through their values," can be interpreted as an endorsement of the substitutional interpretation of higher-order quantifiers as an "intended interpretation." Terms (in particular, variables) for propositional functions occur in argument position to higher-order predicates in the new logic as in the old, so, from a superficial, notational, point of view it seems that propositional functions are thought of as "occurring" *as functions*, and not merely through the occurrence of the particular propositions which are their variables. If the slogan means anything, therefore, it must mean that the truth conditions of sentences containing occurrences of function-terms as arguments must, ultimately, be explicated in terms of sentences in which the function terms occur *as predicates*: sentences, in other words, in which subformulas expressing propositional *values* of the functions occur. Ultimately, on the new view, the only genuine basic facts are such as could be expressed in a First-Order language with appropriate predicates: as Russell more or less says, the role of the machinery of higher-order predication and quantification is simply to abbreviate various infinitary conjunctions and disjunctions of such First-Order statements.

BMT: FORMALISM. We specify the type structure, the vocabulary and formation rules of the formal language, and the deductive apparatus.

Types. There is a BMT-type $\mathfrak{i}$ of individuals. For every non-individual simple type (thus: for the types of propositions and of propositional functions taking arguments

of various simple types) t and every positive integer n, there is a BMT-type t/n of propositional functions (or propositions) of *level n* and *simple type t*. Note that the BMT-type of a non-individual in the ontology of the logic is only partially specified by giving its simple type, but that the single notion of level does the duty of both level and badness in IRT; individuals will again be assumed to be of level 0. Where confusion seems unlikely, "BMT-type" will be abbreviated to "type." The simple type *corresponding* to the BMT-type i is i, and the simple type corresponding to a BMT-type $(\ldots)/n$ is $(\ldots)$.

Non-logical vocabulary. We may have constants denoting individuals, and certain predicate constants. In giving the substitutional interpretation of IRT we assumed that the only primitive predicates were level 1 predicates of individuals, but IRT allows—and in application as an intensional logic demands—predicates of higher type. The intended interpretation of BMT, however, seems to rule such things out. All predicate constants, therefore, are of type $(\mathsf{i}, \ldots, \mathsf{i})/1$ (with 0 or more occurrences of i). For applications it is convenient to assume that one of the predicates expresses identity of individuals, but we shall not include principles of identity in our deductive system.

Logical vocabulary. We have the usual propositional connectives, quantifier symbols, parentheses, brackets, and braces, and many, many variables: for each BMT-type we have alphabets of free and bound variables, and, in addition, for each simple type we have an alphabet of abstract variables (which will only occur bound).

Well-formed expressions. We define the notions of *term* and *formula* by simultaneous recursion:

1. Free variables (i.e., variables from the first class of alphabets) and constants of a given type are *terms* of that type;

2. If F is a term of some type $(t_1, \ldots, t_n)/k$ and $a_1, \ldots, a_n$ are terms with corresponding simple types $t_1, \ldots, t_n$, $F[a_1, \ldots, a_n]$ is a *formula*;

3. Negation and the other truth functional connectives form new *formulas* out of old as usual;

4. If A is a formula, not containing any occurrence of the bound BMT-typed variable x, a is a free variable of the same BMT-type as x, and A^x_a is the result of replacing every occurrence of a in A with an occurrence of x, then $\exists x(A^x_a)$ and $\forall x(A^x_a)$ are both formulas; and

5. If A is a formula, $a_1, \ldots, a_n$ are free variables of BMT-types $t_1, \ldots, t_n$ respectively, $x_1, \ldots, x_n$ abstract variables of the corresponding simple types $t-1, \ldots, t-n$ which do not occur within A, and $A^{a\cdots}_{x\cdots}$ is the result of simultaneously substituting occurrences of $x_1, \ldots, x_n$ for all occurrences of $a_1, \ldots, a_n$ in A, then $\{x_1, \ldots, x_n : A^{a\cdots}_{x\cdots}\}$ is a *term* of the BMT-type $(t-1, \ldots, t-n)/k$, where k is the maximum of the levels of the free BMT-typed variables occurring in $A^{a\cdots}_{x\cdots}$ or $1+$ (the maximum of the levels of the quantified variables occurring

in A), whichever is higher. (Note that this includes the case of *propositional* abstracts: where A is a formula, $\{A\}$ is a term of some type $()/k$.)

Deductive apparatus. The usual machinery of rules for the connectives and for quantifiers of each type (for definiteness we might specify, say, the rules of LK from Gentzen [14]), where the terms substitutable for the bound variables of any given BMT-type are stipulated to be precisely the free variables, constants, and abstracts of that type.

For the purposes of this paper, this completes the description of the logic BMT; axioms of extensionality, of infinity, and for identity will be considered non-logical.

EXTENSIONALITY. The 1925 logic is based on an *extensional* construal of the notion of a propositional function in (at least) one clear sense. Higher-order properties and relations—properties and relations, that is, holding *of* propositions and propositional functions—are countenanced only insofar as they can, in principle, be defined by expressions of an extensional language in which the variables for their arguments occur as (the predicates of) atomic subformulas. No property defined in such a way can distinguish between co-extensive propositional functions. Analogues of set theory's Axioms of Extensionality, therefore, will be valid on Russell's intended interpretation of BMT (though not, of course, derivable from the *logical* system defined in the previous section). The formulation of appropriate extensionality principles, however, is a somewhat delicate matter, and their attribution to Russell somewhat contentious. The problem is that, since the propositional functions of a given simple type come in a topless infinite hierarchy of levels, no quantified variable can range over *all* the potential arguments of a higher-order predicate, making a straightforward statement of extensionality impossible.

In the Introduction to the second edition, Russell (pp. xxxix, xl) somewhat obscurely suggests that certain inferences related to extensionality are legitimate if their premisses are "logical truths," but not if they are only "hypotheses." The idea seems to be something like this. Suppose

$$\forall x(F[x] \leftrightarrow G[x]),$$

where x is a variable for propositional functions of some type, is a *theorem* of the logic. In typical cases it will have been proven by proving

$$(F[a] \leftrightarrow G[a]),$$

with a free variable, a. Now, provided nothing in the proof depended on the level of the variable a, the same proof can be imitated with a higher level variable, and the equivalence thus established for propositional functions of (the same simple type but) arbitrary level. Thus, even though the variable x is of some definite level, the *provability* of

$$\forall x(F[x] \leftrightarrow G[x])$$

often indicates that the equivalence holds at higher levels, whereas its mere truth, or the fact that it follows from assumptions, does not.

Although an extensionality rule of this sort seems to be valid on Russell's intended interpretation of BMT, the sort of extensionality axiom familiar from set theory,

$$\forall F \forall G(\forall x(F[x] \leftrightarrow G[x]) \rightarrow \forall \Phi(\Phi[F] \leftrightarrow \Phi[G])),$$

is deeply problematic. Russell states it (p. xxxix), but it is not quite certain that he meant to endorse it at all types: the variable x was restricted, in an earlier section of the introduction, to ranging over individuals. Now, if, in the displayed formula, the variable x is of type $\mathfrak{i}$ (and F, G, and Φ consequently of types $(\mathrm{i})/i$, $(\mathrm{i})/j$, and $((\mathrm{i}))/k$ for some levels i, j and k), it *is* valid: since there are no distinctions of level among individuals, the antecedent says in full generality that F and G are coextensive, implying that they cannot be distinguished by any Φ definable in an extensional language. If x is of any propositional function type, however, it is not valid. Its implausibility for function types can be seen by considering its corollary,

$$\forall F \forall G(\forall x(F[x] \leftrightarrow G[x]) \rightarrow \forall y(F[y] \leftrightarrow G[y])),$$

where y is of the same simple type as x but of higher level. This says, essentially, that if F and G agree on all functions of some low level, they will *also* agree on all functions of a higher level, but this would only follow if for every higher level function a coextensive function already existed at the lower level. And this is simply Russell's Axiom of Reducibility, which he was not assuming in the introduction and Appendix B, and which is not valid on the intended interpretation of BMT. A counterexample to a simple instance of this form in a specific substitutional model will be presented in section 4D.

SUBSTITUTIONAL SEMANTICS. Let a *substitutional model* be as in the previous section: a structure of the standard sort for the First-Order fragment of the language, with every individual in the domain having a name (so as to avoid complications about satisfaction and assignments). Atomic sentences are true in this model under the usual conditions. We wish to specify what it is for an arbitrary sentence to be true in this model. Truth-functionally compounded sentences depend on their truth-functional constituents in the standard way, and universal (existential) quantifications are true if and only if all (at least one) of their instances are true. If the quantified variable is an individual variable, the instances are the results of deleting the initial quantifier and substituting a name of an individual for all remaining instances of the bound variable. If the quantified variable is of any other BMT-type, the instances are the results of deleting the initial quantifier and substituting some (closed) abstract for all the remaining occurrences of the bound variable (and then, if necessary, reducing to β-normal form). As in the case of IRT, it is intuitively obvious that such a truth definition reduces the truth value of a complex sentence to the truth values of its instances (or truth-functional compounds), and, also as with IRT, it must be shown that the reducibility relation is well-founded: that, in terms of the game metaphor, the game *won't go on forever*. As usual, we show this by defining an ordinal measure of complexity or *grade* to the formulas of the language, and showing that every reduction

of a formula (to an instance, if it is a quantification, or to a truth-functional constituent) leads to a reduction in grade.

The ordinal assignment for formulas of BMT will have to be a bit subtler than that for IRT. We may, as before, assign 0 to atomic formulas having predicates of the language as predicates. With IRT, the grade of an atomic formula with a variable as its predicate was determined by the *badness* of the predicate, but this will not do for BMT: since, in BMT, arguments may have the same level as, or a higher level than, the predicate of an atomic formula, the levels of the arguments have to be taken into consideration as well: consider the atomic formula

$$\Phi[\{A\}]$$

where Φ is a variable of type $(())/1$, and A is a formula containing variables of high level. Admissible substituents of Φ will include

$$\{p : p\}$$

and

$$\{p : (p \ \& \ p)\},$$

so the atomic formula will have substitution instances which (after β-reduction) include A itself and the conjunction $(A \ \& \ A)$. The grade of the atomic formula, therefore, will have to reflect the complexity of the argument. It turns out that the appropriate strategy is to make the grade of an atomic formula depend on the *maximum* of the levels of its terms, regardless of whether the maximum is achieved by the predicate or an argument (or both). Where one or more of the arguments are *abstracts*, they will be deemed to have the same level as the lowest-level BMT-typed variables substitutable for them: the level of an abstract is therefore that of the highest-level variables occurring free anywhere within it, *including occurrences in smaller abstracts occurring in it*, or one more than the highest-level quantified variable occurring anywhere within it, *including occurrences in smaller abstracts*, whichever is higher.

In defining the grades of more complex formulas for IRT, we were able to use a particularly simple scheme for the purposes of the substitutional semantics, where only closed formulas (sentences) were at issue. Because the open formulas substitutable for a variable could contain free variables of the same badness as the first variable itself, however, a more complicated scheme was needed for proof-theoretic purposes. Since the term of an atomic formula with highest level in BMT may not be the predicate, the sort of problem that forced us to use the more complex scheme with open formulas in IRT arises in BMT arises even in the semantics of BMT: A, in the previous example, can be chosen to be a sentence. Again, in IRT we were able to keep the ordinals low (the limit of the grades of formulas of IRT is ω^2) by the trick of raising the grade by only 1 for an *outer* quantifier with a variable of a certain badness, once an inner quantifier with a variable of the same or worse badness had been recorded. This, however, was possible only because the arguments of an atomic formula are not as bad as the predicate, so that when a predicate variable is eliminated, the formula that

replaces an atomic formula with the given variable as predicate must be built up from atomic formulas of strictly lower grade. In BMT the level of a quantified variable must be recorded in the grade of a formula regardless of how high the grade of the matrix of the quantification is. Thus, for example, let Φ in our previous example of a formula be of type $(())/2$, with A containing bound variables of higher level. Then, if we tried to use the trick that kept the grades low in IRT, our grade for

$$\exists\Phi(\Phi[\{A\}])$$

would be (level of $\{A\}$) + 1. One admissible substituent for a variable of type $(())/2$, however, is

$$\{p : \exists\Psi(\Psi[p])\},$$

where Ψ is a variable of type $(())/1$, so that

$$\exists\Psi(\Psi[\{A\}])$$

turns out (after β-reduction) to be an instance of

$$\exists\Phi(\Phi[\{A\}]).$$

Clearly, the grade of a formula of this form must record the level of the variable quantified by the initial quantifier.

A definition of grade satisfying these desiderata may be defined as follows:

- the *grade* of an atomic sentence or quasi-formula with a predicate constant as its predicate is 0
- the grade of an atomic quasi-formula with a predicate variable as its predicate is $\omega^{(2k-1)}$, where k is the level of that variable, or of the highest-level quasi-term occurring as an argument to it, whichever is higher
- the grade of a negation is the grade of the negated formula (or quasi-formula) + 1
- the grade of a conjunction (etc) is the grade of whichever conjunct (...) has the higher grade + 1
- the grade of a quantification with an individual variable is the grade of the matrix + 1
- the grade of a quantification with a higher-order variable of level k is the grade of the matrix + ω^{2k}.

(Note that the desire to record the level of variables of quantification independently of the grade of the matrix forces us to make our "jumps" by exponentiation rather than, as with the grade definition for IRT, multiplication. As a result the limit of the grades of BMT formulas is ω^{ω} instead of ω^2.)

It remains to verify that a reduction of a formula always reduces grade. The cases of truth-functional compounds and quantifications with individual variables are

standard, leaving the case of a higher-order quantification with a variable of level k. The initial effect of deleting the quantifier is to subtract ω^{2k} from the grade: in terms of the standard notation for ordinals, to delete a final addend of the form ω^{2k} (or to reduce by one the parameter in a final addend of the form $(\omega^{2k}) \cdot n$). The grade of the matrix may increase, but not by enough to undo the effect of this subtraction. Note first that the substitution of an abstract for a variable *as an argument* in an atomic (quasi-)formula cannot increase the grade of that atom, so we only need to worry about substitutions of abstracts for predicates in atomic quasi-formulas. An abstract substitutable for a bound variable of level k may contain bound variables of level no higher than $k - 1$, and may also (in the proof-theoretic environment: this does not arise in the substitutional semantics, in which only the substituents considered are *closed* abstracts) contain free variables of level no higher than k. Concretizing to get rid of the abstract, we get a (quasi-)formula built up from atoms having, as their predicates and arguments, free and bound variables of the abstract and/or argument terms from the original atom. (If some of the arguments of the original atom were themselves abstracts, further concretion may be needed to obtain a β-normal form, yielding additional possibilities for the atoms of the new (quasi-)formula. Since, however, the grade of the original atom already reflected levels of the highest-level variables occurring in it, including occurrences within abstracts, it can be seen that this further complication does not affect the argument.) Thus the highest possible grade for an atomic (quasi-)subformula of the new (quasi-)formula is the same as the grade of the atom we started with. The new (quasi-)formula is built up from these atoms by means of truth-functional connectives and quantifiers with variables of level no higher than $k - 1$. Its grade, therefore, will be less than the grade of the original atom $+ \omega^{(2k-1)}$. Substitution of formulas of this grade for occurrences of the original atom, in turn, can raise the grade of the matrix of the original quantification by no more than $\omega^{(2k-1)}$, so the reduction has yielded a formula of lower grade.

Note that this notion of grade can be used to both to establish the termination of the semantic "game" (and so, as argued, the philosophical correctness of the formal logic BMT given Russell's "constructivistic" intentions) and also to prove the Hauptsatz for BMT.

4. Mathematical Induction and Landini

In this section we discuss (first subsection) the general question of the utility of ramified systems for the foundations of mathematics, (second) Russell's Appendix B, (third) Landini's claim, and (fourth) make good on our claim from the previous section that Landini's extensionality axiom scheme is invalid on the substitutional semantics.

ARITHMETIC IN RAMIFIED LOGICS. The first edition of *Principia* postulated reducibility. Given infinitely many individuals this allows the definition of cardinal numbers as ((i))s (the equivalence classes of (i)s under the equipollence relation) and

the definition of the finite or natural numbers as those belonging to every (((ι))) containing 0 and closed under successor. First- and Higher-Order Peano Arithmetics are interpretable in the system, with the rule of mathematical induction (IND) derived by simple universal quantifier elimination from the definition of natural number. The Axiom of Reducibility, however, is introduced somewhat apologetically, as justified mainly in terms of its consequences. What could Russell have done if he had combined *his* philosophy of logic, including a certain scepticism about Reducibility, with Brouwer's (or Weyl's) personality and willingness to advocate trimming mathematics to eliminate the philosophically suspect bits? More precisely, how much of this development is possible in a Ramified system: say, IRT plus a weak Axiom of Infinity? For simplicity, assume that identity of individuals is given as a primitive predicate constant of type (ι, ι)/1. (This assumption can be avoided—recall that the system with identity can be interpreted in that without—at the expense of using variables a level of badness higher in our definitions.)

It turns out that the definitions of cardinal number and the most familiar arithmetic functions (addition, multiplication and—with a bit more work—exponentiation) can be taken over with minimal changes, and, when *finite* numbers have been appropriately defined, can be proven to be total and to satisfy their usual recursion equations over the finite numbers. Additional functions can be defined by composition and by *bounded* primitive recursion. The situation with IND, however, is delicate. The quantification over (((ι)))s has to be restricted to some level or other (as well as having the arguments specified as to level!)—for example, to (((ι)/1)/1)/3s. The derivation of IND then applies only to conditions which define propositional functions of this type: in particular, to conditions which, when written out in primitive notation, contain no quantified variables of this type. Most of the induction axioms of (First-Order) Peano Arithmetic are therefore, not derivable in IRT + Infinity: quantification over natural numbers is represented by quantification over cardinals *restricted* to finite ones, so any number-theoretic condition containing (unbounded) quantification over natural numbers translates into one containing forbidden bound variables! It is, however, possible to prove that numbers less than a given natural number are themselves natural, so that IND for number-theoretic conditions containing only *bounded* numerical quantifiers *is* available. (Ramified Fourth-Order Logic is more complicated than most mathematical logicians want their underlying systems to be. A small fragment of arithmetic is derived from an infinity axiom in Ramified Second-Order Logic, using "logicist" definitions in the spirit of Frege and Russell, in [21]. [5] makes further simplifications and—in sections due solely to Burgess—establishes upper and lower bounds on how much arithmetic is obtainable on this basis.)

APPENDIX B. BMT is a proper supersystem of IRT, so it is not immediately apparent that a foundational system employing it will have the same mathematical weakness as IRT + Infinity. In the second of the three appendices added to the (first volume of the) second edition of *Principia*, Russell claimed to have derived unrestricted IND in BMT + Infinity. Had the claim been correct, this system would have served as a natural foundation for *predicative analysis*, as advocated by Weyl and his followers.

(Since BMT is a weaker Higher-Order Logic than Simple Type Theory, this would have meant that the whole of First-Order Peano Arithmetic was derivable, but not Higher-Order Arithmetics: in the Introduction to the second edition Russell notes that BMT does not yield classical analysis.) The claim is, however, wrong: the proof given is irremediably flawed, as Gödel [16] appears to have been the first to note.

Still, there is one point in the supposed proof suggesting that BMT *might* have interesting mathematical advantages over IRT. Lemma *89.12 states that every propositional function true of only finitely many objects is coextensive with a minimal level propositional function. This lemma is *correct*, and of considerable use in deriving mathematics within ramified systems like IRT and BMT. (For examples of its use, see [21].) It is provable in a stronger form in BMT, however, than in IRT. For simplicity let us consider propositional functions true of individuals and assume that identity of individuals is expressed by a primitive of type (i, i)/1. For convenience we adopt set-theoretic vocabulary ("subset," "member," ...) for talking about propositional functions and the items of which they are true. Recall also that levels are cumulative, so that (i)/1 "sets" are included among the (i)/2s.

In IRT we may define, for example, an (i)/2 set, X, to be *finite* if and only if it belongs to every ((i)/2)/1 class which (a) includes all unit subsets of X, and (b) for each proper (i)/2 subset, Y, of X included and each individual, x, in $X - Y$ includes also the union of Y with the unit set of x. (A fairly standard set-theoretic definition of finite set.) The class of (i)/2 subsets of X which are coextensive with (i)/1 sets is a ((i)/2)/1 class, denoted by the abstract

$$\{Y : \forall x(Y[x] \rightarrow X[x]) \ \& \ \exists A \forall x(Y[x] \leftrightarrow A[x])\},$$

where X and Y are of type (i)/2, A of type (i)/1, and x of type i. We now prove that it fulfills conditions (a) and (b).

(a) Let Y be a unit (i)/2 subset of X and x be its member. Then

$$\{y : y = x\}$$

is a (i)/1 set coextensive with Y.

(b) Let Y be an (i)/2 subset of X, let x be a member of X not in Y, and let A be an (i)/1 set coextensive with Y. Then

$$\{y : A[y] \vee y = x\}$$

is an (i)/1 set coextensive with the (i)/2 set

$$\{y : Y[y] \vee y = x\}.$$

This proves a strictly limited version of *89.12: all finite (i)/2 sets are coextensive with (i)/1 sets, but nothing is said about, say, finite (i)/3 sets: since they cannot belong to ((i)/2)/1 classes, the definition of finiteness does not apply to (i)/3 sets.

In BMT, by contrast, *89.12 can be proven in greater generality. The quantified class variable in the definition of finiteness may be taken to be of BMT-type ((i))/2 (or

of higher level with the same simple type), capable of taking as arguments variables of simple type (i) and *any* level, and the bound set variables in conditions (a) and (b) as of BMT-type (i)/2. The definition is applicable to sets of type (i)/n for any n, and any finite set of any such BMT-type can be shown, by the same argument as before, to be coextensive with an (i)/1.

Unfortunately for Russell, the result that he *really* needed, that every *subset* of a finite set is finite (his *89.17) is still, apparently, not forthcoming. Leaving the question of just what the mathematical strength of BMT + Infinity *is*. Gödel leaves the question open:

> So the question whether (or to what extent) the theory of integers can be obtained on the basis of the ramified hierarchy must be considered as unsolved at the present time. ([16]: 146)

Burgess's negative result in [5] for a system based on Ramified Second-Order Logic is established by reference to Gödel's Incompleteness Theorem: the system cannot yield the mathematics needed for a proof of its own consistency, where a key step in the obvious consistency proof is the establishment of Gentzen's *Hauptsatz* for the logic. In this connection, the fact that the complexity ordering of sentences used in establishing the substitutional semantics and *Hauptsatz* for BMT is appreciably longer than that needed for IRT suggests that it may require stronger assumptions for its proof than does that for IRT, but it would certainly seem to provable in, say, (First-Order) Peano Arithmetic. Given the *Hauptsatz* for the logic of BMT, a consistency proof for BMT + Infinity (with identity for individuals primitive and axioms of substitution of identicals for all propositional functions taking individuals as arguments) is straightforward. By the *Hauptsatz*, any First-Order conclusion derivable from First-Order premisses in BMT is derivable from them in First-Order Logic: any BMT-provable sequent containing only First-Order formulas in the antecedent and succedent is a valid sequent of First-Order Logic. To prove the consistency of BMT + Infinity, then, it suffices to give an evidently consistent First-Order theory from which the axiom of Infinity is derivable in BMT. Consider the (decidable) theory of dense linear order (with, for definiteness, no top or bottom), formulated with identity and the order relations as primitives. (To turn this example into a concrete example of a substitutional model, take the rational numbers as the domain of individuals, with names given by appending fractions as subscripts to some constant symbol.) General substitution of identicals is not derivable from the First-Order axioms for identity: substitution axioms allowing the substitution of terms for identicals as arguments to the predicate constants of the First-Order language will not yield substitution as arguments to (i)/1 predicate variables. The system with general substitution of (individual) identicals is, however, interpretable in the system without: simply restrict all higher-order quantifiers of appropriate types to propositional functions for which substitution holds. A good axiom of Infinity asserts the existence of a ((i))/1 class of non-empty (i) sets containing a proper superset of each of its members. But the class of sets of the form

$$\{x : a < x \;\&\; x < b\}$$

for $a < b$ (these are all, for what it's worth, (i)/1 sets) will work, and can be proven to work. Given the elementary nature of the consistency proof just sketched, it seems very unlikely that full IND (even for conditions in the language of First-Order Peano Arithmetic) is derivable in BMT + Infinity, but determining exactly what subsystem of PA *is* obtainable would require a more refined analysis. It seems that essentially the argument given for *89.12 allows the "Pigeon-Hole" principle to be proven in a much more general form in BMT + Infinity than in IRT + Infinity, so the 1925 logic may be of greater mathematical interest than has been believed!

LANDINI. In [25] Landini attempts to vindicate Russell by showing that the derivation of Appendix B *could* have been carried out successfully after all. (Landini is concerned with provability and not with definability; the potentially misleading title of his article is a reference to [32].) What he presents is a derivation of the key claim of Appendix B, *89.17, from BMT + Infinity … *plus* Extensionality. Landini takes the extensionality axiom discussed in section 3C above to hold for propositional functions in general, and not just for functions taking only individuals as arguments. Landini's derivation is correct, and—Russell's exposition not being as explicit as one could desire—it may be that Russell, in 1925, believed that the extensionality axiom held on the "constructive" interpretation motivating BMT. If he did believe this, he was wrong, as we demonstrate in the next subsection. The logic BMT is primarily of interest in virtue of having this constructivistic philosophical motivation, and since the extensionality principle Landini needs is not valid on this interpretation, his derivation does not really vindicate Russell.

INVALIDITY OF LANDINI'S AXIOM. In his replacement for Russell's Appendix B in [25], Landini appeals to a general extensionality principal which we have already suggested is incompatible with the philosophical motivation for BMT. Subject to reservations already mentioned, we may take the interpretation on which the higher-order quantifications are interpreted substitutionally as a reasonable approximation to the intended interpretation. If we can display a reasonable substitutional model on which instances of the extensionality principal come out false, therefore, we will have shown that Landini's axiom is unacceptable.

Consider, therefore, the substitutional model built up over standard Arithmetic: let the individuals be just the natural numbers, and take as primitive vocabulary just the usual vocabulary of First-Order Peano Arithmetic (reformulated, given our "official" syntax, with predicates replacing the usual function symbols). The range of the type (i)/1 variables on a substitutional interpretation being precisely those subsets of the individual domain as are (parametrically) definable by First-Order formulas of the language, the (i)/1 propositional functions on this interpretation correspond to the *arithmetic* sets of natural numbers, and so on.

(Note that we are *not* now trying to "construct" arithmetic on a logicist basis, but are rather *assuming* arithmetic in a metamathematical investigation of a logicist logic, BMT: even though our goal is to evaluate Landini's claims of success at logicist construction, there is no circularity in our assuming arithmetic.)

The relevant extensionality principle is the schema

$$\forall X(A(X) \leftrightarrow B(X)) \rightarrow \forall F(A(F) \leftrightarrow B(F)),$$

where X and F may be taken to be of types (i)/1 and (i)/2 respectively, and A and B are arbitrary appropriate contexts. (This schema is stated, in full generality, at the beginning of section 4 of [25]; it is essential to Landini's proof.) In the particular substitutional model specified, then, X ranges over arithmetic sets of numbers. By Tarski's Theorem, the set of (Gödel numbers of) true sentences of First-Order Arithmetic is *not* in the range of X. On the other hand, it is fairly easy to specify, in First-Order Arithmetic supplemented with set variables, the conditions a set must satisfy in order to *be* the set of truths of First-Order Arithmetic (for details, see Chapter 19 of [4]—the claim is that the *unit class* of the set of truths of (First-Order) Arithmetic *is* definable in Arithmetic). We may, therefore, choose $A(\cdot)$ to express (on the specified interpretation) $(\cdot)$ *is the set of Arithmetic truths* and (there being no difficulty at all in expressing being-the-negation-of *via* Gödel numbering) choose $B(\cdot)$ to express $(\cdot)$ *is the set of Arithmetic falsehoods*. The antecedent, $\forall X(A(X) \leftrightarrow B(X))$, of this instance of the schema is, then, vacuously true: no item in the range of X satisfies either $A(\cdot)$ or $B(\cdot)$.

Now, it is also possible, for any given natural number n, to define in First-Order Arithmetic the set of (Gödel numbers of) truths of First-Order Arithmetic containing n or fewer quantifiers (details, again, in [4]): all these sets are, on the specified substitutional model, (i)/1s. It is, moreover, possible to write out a condition (with no quantified set variables) specifying exactly these "partial truth sets." In the specified substitutional model of BMT, to put it in other words, the class of partial truth sets is defined by a propositional function of type ((i))/1. But a sentence is a truth of Arithmetic if and only if it is a member of some partial truth set! Quantifying a variable of type (i)/1, then, we can write a formula expressing $(\cdot)$ *is a truth of First-Order Arithmetic*. The set of truths of First-Order Arithmetic, then, amounts to a (i)/2, as is, of course, the set of falsehoods. The set of truths is not identical to the set of falsehoods, so the consequent, $\forall F(A(F) \leftrightarrow B(F))$, is false. Q.E.D.

Acknowledgement. An earlier, and drastically condensed, version of this paper has been published as [23]. While accepting full responsibility for the present paper, I would like to thank Dr. Davoren for her essential contribution to earlier phases of the research.

References

[1] Anderson, C. Anthony: 1986. Some difficulties concerning Russellian intensional logic. *Noûs* 20: 35–43.

[2] Anderson, C. Anthony: 1989. Russellian intensional logic. In: J. Almog *et al.* (eds.), *Themes from Kaplan*, New York: Oxford University Press.

[3] Benacerraf, Paul, and Hilary Putnam: 1964. *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs: Prentice-Hall.

[4] Boolos, George, and Richard Jeffrey: 1974. *Computability and Logic*. Cambridge, UK: Cambridge University Press. Later editions 1980 and 1989.

[5] Burgess, John P., and Allen P. Hazen: 1999. Predicative logic and formal arithmetic. *Notre Dame Journal of Formal Logic* 39 (1998 cover date): 1–17.

[6] Carnap, Rudolf: 1928. *Der Logische Aufbau der Welt*. Berlin: Weltkreis. Eng. tr. *The Logical Structure of the World*, Berkeley & Los Angeles: University of California Press, 1967.

[7] Church, Alonzo: 1956. *Introduction to Mathematical Logic*. Princeton: Princeton University Press.

[8] Church, Alonzo: 1974. Russellian simple type theory. *Proceedings and Addresses of the American Philosophical Association* 47: 21–33.

[9] Church, Alonzo: 1976. Comparison of Russell's resolution of the semantical antinomies with that of Tarski. *Journal of Symbolic Logic* 41: 747–760. Repr. with correction in R. L. Martin (ed.), *New Essays on Truth and the Liar Paradox*, New York: Oxford University Press.

[10] Church, Alonzo: 1984. Russell's theory of identity of propositions. *Philosophia Naturalis* 21: 513–522.

[11] Davoren, Jen M. and Allen P. Hazen: 1991. Russell, Gödel and Skolem: how much of arithmetic is predicative? *Journal of Symbolic Logic* 56: 1108–1109.

[12] Dunn, John M. and Nuel D. Belnap: 1968. The substitution interpretation of the quantifiers. *Noûs* 2: 177–185.

[13] Fitch, Frederick B.: 1938. The consistency of the ramified *Principia*. *Journal of Symbolic* Logic 3: 140–149.

[14] Gentzen, Gerhard: 1934. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39: 176–210 and 405–431. Eng. tr.: Investigations into logical deduction, *American Philosophical Quarterly* 1 (1964): 288–306 and 2 (1965): 204–218. Eng. tr. repr. in [15].

[15] Gentzen, Gerhard: 1969. *Collected Papers of Gerhart Gentzen*. Edited by M. Szabo, Amsterdam: North-Holland.

[16] Gödel, Kurt: 1944. Russell's mathematical logic. In P. A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, Evanston: Northwestern University. Repr. in [17] and other places.

[17] Gödel, Kurt: 1990. *Collected Works*, vol. II. Edited by S. Feferman *et al*., New York: Oxford University Press.

[18] Goodman, Nelson: 1951. *The Structure of Appearance*. Cambridge, MA: Harvard University Press.

[19] Grover, Dorothy: 1973. Propositional quantification and quotation contexts. In [27].

[20] Hazen, Allen P.: 1985. Nominalism and abstract entities. *Analysis* 45: 65–68.

[21] Hazen, Allen P.: 1992. Interpretability of Robinson's arithmetic in the Ramified Second-Order theory of dense linear order. *Notre Dame Journal of Formal Logic* 33: 101–111.

[22] Hazen, Allen P.: 1995. On quantifying out. *Journal of Philosophical Logic* 24: 291–319.

[23] Hazen, Allen P. and Jen M. Davoren: 2000. Russell's 1925 Logic. *Australasian Journal of Philosophy* 78: 534–556.

[24] Henkin, Leon: 1950. Completeness in the theory of types. *Journal of Symbolic Logic* 15: 81–91. Repr. in J. Hintikka (ed.), *Philosophy of Mathematics*, London: Oxford University Press, 1969.

[25] Landini, Gregory: 1996. The definability of the set of natural numbers in the 1925 *Principia Mathematica*. *Journal of Philosophical Logic* 25: 597–615.

[26] Landini, Gregory: 1998: *Russell's Hidden Substitutional Theory*. New York: Oxford University Press.

[27] Leblanc, Hugues (ed.): 1973. *Truth, Syntax and Modality*. Amsterdam: North-Holland.

[28] Leblanc, Hugues: 1975. That *Principia Mathematica*, first edition, is predicative after all. *Journal of Philosophical Logic* 4: 67–70.

[29] Leblanc, Hugues, and George Weaver: 1973. Truth-functionality and the ramified theory of types. In [27].

[30] Linsky, Bernard: 1999. *Russell's Metaphysical Logic*. Stanford: CSLI.

[31] Lorenzen, Paul: 1951. Algebraische und logistische Untersuchungen über freie Verbände. *Journal of Symbolic Logic* 16: 81–106.

[32] Myhill, John: 1974. The undefinability of the set of natural numbers in the ramified *Principia*. In: G. Nakhnikian (ed.), *Bertrand Russell's Philosophy*, New York: Harper and Row.

[33] Myhill, John: 1979. A refutation of an unjustified attack on the axiom of reducibility. In: G. W. Roberts (ed.), *Bertrand Russell Memorial Volume*, London: George Allen & Unwin.

[34] Parsons, Charles: 1971. Mathematics and ontology. *Philosophical Review* 80: 151–176. Repr. in [38].

[35] Parsons, Charles: 1971. A plea for substitutional quantification. *Journal of Philosophy* 68: 231–237. Repr. with added note in [38].

[36] Parsons, Charles: 1974. Sets and classes. *Noûs* 8: 1–12. Repr. in [38].

[37] Parsons, Charles: 1982. Substituional quantification and mathematics (review of Gottlieb, *Ontological Economy*). *British Journal for the Philosophy of Science* 33: 409–421.

[38] Parsons, Charles: 1983. *Mathematics in Philosophy*. Ithaca: Cornell University Press.

[39] Pears, David: 1972. *Bertrand Russell: a Collection of Critical Essays*. New York: Doubleday.

[40] Quine, Willard V.: 1960. *Word and Object*. Cambridge: MIT Press.

[41] Quine, Willard V.: 1963. *Set Theory and its Logic*. Cambridge: Harvard University Press.

[42] Ramsey, Frank P.: 1925. The foundations of mathematics. *Proceedings of the London Mathematical Society* 25: 338–384. Repr. in [43, 44, 45].

[43] Ramsey, Frank P.: 1931. *Foundations of Mathematics and Other Logical Essays*. Edited by R. B. Braithwaite, London: Routledge and Kegan Paul.

[44] Ramsey, Frank P.: 1978. *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. Edited by D. H. Mellor, London: Routledge and Kegan Paul.

[45] Ramsey, Frank P.: 1990. *Philosophical Papers*. Edited by D. H. Mellor, Cambridge, UK: Cambridge University Press.

[46] Russell, Bertrand: 1903. *The Principles of Mathematics*. London: Cambridge University Press. Second edition with new introduction London: Allen & Unwin, 1925.

[47] Russell, Bertrand: 1910–1911. On knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society* 11: 108–128. Repr. in [48].

[48] Russell, Bertrand: 1918. *Mysticism and Logic*. London: Longmans, Green.

[49] Russell, Bertrand: 1924. Logical atomism. In: J. H. Muirhead (ed.), *Contemporary British Philosophy* (First Series), London: George Allen & Unwin. Repr. in [51].

[50] Russell, Bertrand: 1925. *Introduction* (pp. xiii–xlvi) and three Appendices (pp. 635–666) added to the second edition of [57]. The Introduction and Appendices A and C (pp. 401–408) (but not B) are included in [58].

[51] Russell, Bertrand: 1956. *Logic and Knowledge*. Edited by R. C. Marsh, London: George Allen & Unwin.

[52] Schütte, Kurt: 1960. *Beweistheorie*. Berlin: Springer-Verlag.

[53] Skolem, Thoralf: 1962. *Abstract Set Theory.* Notre Dame Mathematical Lectures 8. Notre Dame: Notre Dame University.

[54] Taylor, Barry M. and Allen P. Hazen: 1995. Flexibly structured predication. *Logique et Analyse* 139–140 (1992 cover date): 375–393.

[55] Toledo, Sue: 1975. *Tableau Systems*. Lecture Notes in Mathematics, vol. 447. Berlin: Springer.

[56] Weyl, Hermann: 1918. *Das Kontinuum*. Leipzig: De Guyter. Eng. tr. *The Continuum*, Kirksville, Missouri: Thomas Jefferson University Press, 1987.

[57] Whitehead, Alfred N. and Bertrand Russell: 1910. *Principia Mathematica*, vol. I. Cambridge, UK: Cambridge University Press. Second edition with additional matter by Russell, 1925.

[58] Whitehead, Alfred N. and Bertrand Russell: 1962. *Principia Mathematica* to *56. Abridged reprint of second edition of [1910]. Cambridge, UK: Cambridge University Press.

Philosophy Department
University of Melbourne
Victoria 3010
Australia

E-mail: a.hazen@philosophy.unimelb.edu.au

Russell on Method

Andrew D. Irvine

Abstract. This paper explores a tension between Russell's views on epistemology and his views on philosophical method. According to Russell-the-foundationalist, epistemic certainty is a common feature of both logical and empirical knowledge. According to Russell-the-fallibilist, the analytic method implies a universal lack of certainty, not just in philosophy, but everywhere. It is argued that it was this tension between epistemology and method that eventually resulted in Russell abandoning his foundationalism in favour of an alternative theory of knowledge.

Bertrand Russell's quest for certainty is both famous and well documented. "I wanted certainty in the kind of way in which people want religious faith," he says in his autobiography. (See [39]: vol. 3, 220) Understood in this way, his foundationalism in epistemology, while in some respects groundbreaking, is fully in keeping with traditional epistemic goals. For Russell-the-foundationalist, the aim of epistemic justification is epistemic certainty.

In contrast, Russell's views about method stress the uncertainty of all logical, philosophical and scientific claims. In his many writings about what he calls the "liberal" or "scientific" outlook, Russell-the-fallibilist continually stresses the uncertainty of all branches of justified belief. In these writings, Russell-the-foundationalist is nowhere to be seen.

This apparent tension between Russell's views on epistemology and his views on method is significant. According to Russell-the-foundationalist, epistemic certainty is a common feature of both logical and empirical knowledge. In contrast, according to Russell-the-fallibilist, both the analytic method and the liberal or scientific outlook essentially imply a universal lack of certainty, not just in philosophy, but everywhere. In what follows, it is argued that it was this tension between epistemology and method that eventually resulted in Russell abandoning his foundationalism in favour of an alternative theory of knowledge.

This paper is divided into four main parts. The first reviews Russell's foundationalism with regard to both logical and empirical knowledge. The second reviews Russell's theory of analysis and his views about method more generally. The third investigates the tension between Russell's theory of knowledge and his theory of method, and concludes that this tension needs to be resolved. The fourth describes Russell's resolution. The paper finishes with some remarks about the significance of this issue, not only for Russell, but for analytic philosophy more generally.

1. Russell's Theory of Knowledge

Here is Russell at age 87 reviewing his life's work:

> There is only one constant preoccupation: I have throughout been anxious to discover how much we can be said to know and with what degree of certainty or doubtfulness. ([37]: 9)

Here he is making much the same point thirty-five years earlier:

> From early youth, I had an ardent desire to believe that there can be such a thing as knowledge, combined with a great difficulty in accepting much that passes as knowledge. It seemed clear that the best chance of finding indubitable truth would be in pure mathematics … . ([31]: 323)[1]

In other words, throughout his life, Russell's desire for certainty permeated his philosophy. As A. J. Ayer puts it, Russell "was a consistent sceptic, in the sense of holding that all our accepted beliefs are open to question; he conceived it to be the business of philosophy to try to set these doubts at rest." ([1]: 66)

The main reason Russell believes that most propositions are open to doubt is that they refer to entities whose existence is questionable. Because of this, the best way to resolve such doubts is to identify or reduce each class of questionable entities to a separate class of entities whose existence is more certain.

Russell observes that most of our beliefs involve inferences of one kind or another. For example, unlike his idealist opponent, the early Russell accepts independently existing physical objects as the underlying cause of perceptual experience.[2] His common-sense justification of these objects uses what we would today call "inference to the best explanation." By postulating independently existing physical objects, we are better able to explain recurring patterns of experience. In short, we infer the existence of such objects from their explanatory power.

This type of inference originally played a fundamental role in Russell's philosophy, especially in light of worries about naïve realism. According to Russell, not only the argument from illusion, but science itself makes it difficult to accept the claim that we directly or naïvely perceive physical objects. As he puts it in *An Inquiry into Meaning and Truth*, "Naïve realism leads to physics, and physics, if true, shows that naïve realism is false. Therefore, naïve realism, if true, is false: therefore it is false." ([33]: 15)[3]

Despite this, Russell also points out that

[1] Or as he puts it in [37]: 154–155, "My philosophical development, since the early years of the present century, may be broadly described as a gradual retreat from Pythagoras. … I hoped, at that time, that all science could become mathematical, including psychology. … My interest in the applications of mathematics was gradually replaced by an interest in the principles upon which mathematics is based. This change came about through a wish to refute mathematical scepticism."

[2] See [22]: ch. 2. Russell later changes his mind on this issue. For example, in his [24] he attempts to reduce physical objects to collections of sense-data. However, by the time he is writing both his [32] and his [36] he has returned to the view that physical objects are best understood as inferred entities.

[3] It was this encapsulation of Russell's argument against naïve realism that impressed Einstein.

> our world is not wholly a matter of inference. There are things that we know without asking the opinion of men of science. If you are too hot or too cold, you can be perfectly aware of this fact without asking the physicist what heat and cold consist of. … We may give the name 'data' to all the things of which we are aware without inference. ([37]: 17)

We are thus led to Russell's distinction between two kinds of knowledge of truths: that which is direct, intuitive, certain and infallible, and that which is indirect, derivative, uncertain and open to error.[4] To be justified, every indirect knowledge claim must be validated by showing how it is derived from more fundamental, direct or intuitive knowledge. The kinds of truths that are capable of being known directly include both truths about immediate facts of sensation and truths of logic.[5]

Eventually, Russell supplemented this distinction between direct and indirect knowledge with his famous distinction between knowledge by acquaintance and knowledge by description.

As Russell puts it, "I say that I am *acquainted* with an object when I have a direct cognitive relation to that object, i.e., when I am directly aware of the object itself. When I speak of a cognitive relation here, I do not mean the sort of relation which constitutes judgment, but the sort which constitutes presentation." ([21]: 209) Later, he clarifies this point by adding that acquaintance involves, not knowledge of truths, but knowledge of things, cf. [22]: 44. Thus, while intuitive knowledge and derivative knowledge both involve knowledge of propositions (or truths), knowledge by acquaintance and knowledge by description both involve knowledge of objects (or things).[6] Since it is those objects with which we have direct acquaintance that are the least questionable members of our ontology, it is these objects upon which Russell ultimately bases his epistemology.

A paradigmatic example of the reduction of one questionable class of entities to another, less questionable one, involves the reduction of material objects to sense-data. For example, in *Our Knowledge of the External World*, Russell explains that

> A thing may be defined as a certain series of appearances, connected with each other by continuity and by certain causal laws. … More generally, a 'thing' will be defined as a certain series of aspects, namely those which would commonly be said to be *of* the thing. To say that a certain aspect is an aspect *of* a certain thing will merely mean that it is one of those which, taken serially, *are* the thing. ([24]: 106–107)

Another example involves the reduction of numbers to classes. As Russell puts it

[4] For example, see [19]: 41f, as well as [21], [22], and [25].

[5] For example, see [22]: 109 where he states that propositions with the highest degree of self-evidence (what he here calls "intuitive knowledge") include "those which merely state what is given in sense, and also certain abstract logical and arithmetical principles, and (though with less certainty) some ethical propositions."

[6] This distinction is slightly complicated by the fact that, even though knowledge by description is in part based upon knowledge of truths, it is still knowledge of things, and not of truths. I am grateful to Russell Wahl for reminding me of this point.

> Two equally numerous collections appear to have something in common: this something is supposed to be their cardinal number. But so long as the cardinal number is inferred from the collections, not constructed in terms of them, its existence must remain in doubt. … By defining the cardinal number of a given collection as the class of all equally numerous collections, we … thereby remove a needless doubt from the philosophy of arithmetic. ([26]: 156)

In much the same way, points and instants are reduced to ordered classes of volumes and events, and classes themselves are reduced to propositional functions.

To obtain these kinds of reductions, Russell requires that we adopt what he calls "the supreme maxim in scientific philosophizing," namely the principle that "Whenever possible, logical constructions," or as he also sometimes puts it, logical fictions, "are to be substituted for inferred entities."[7] Ontological atoms turn out to be anything that is not a construction in this sense. They are the objects that resist reduction to other objects. Such objects are atomic, both in the sense that they fail to be composed of individual, substantial parts, and in the sense that they exist independently of one another. Their corresponding propositions are also atomic, both in the sense that they contain no other propositions as parts, and in the sense that the members of any pair of true atomic propositions will be logically independent of one another. It turns out that formal logic, if carefully developed, will provide a mechanism for mirroring precisely, not only the various relations between all such propositions, but their various internal structures as well.

In one sense this goal of having logic mimic reality originates with Plato, for it was Plato who first enjoined us to carve nature at its joints.[8] However, in a more precise sense, it originates with Leibniz, for it was Leibniz who first championed the use of a logically ideal language, together with an instrument of universal deductive reasoning. Together, Leibniz's *characteristica universalis* and *calculus ratiocinator* were intended to help us represent the world as it actually is and to reason about it without fear of error. As Russell himself tells us, Leibniz believed that, with these tools, "we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis … If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other (with a friend as witness, if they liked): Let us calculate." ([17]: 169–170) Thus, like Leibniz, Russell held that advances in metaphysics would occur only when accompanied by advances in logic.

Even so, at this point we are entitled to inquire further about the nature of Russell's ontological atoms. How is it that they are to be identified?

---

[7] See [26]: 155. Cf., too, [24]: 107, and [31]: 326.

[8] For example, see Plato's *Statesman* in [11]: 287c, where it is suggested that conceptual divisions should be made "according to their natural divisions as we would carve a sacrificial victim."

Russell's answer is clearly influenced by his foundationalism. For Russell, genuine ontological atoms are exactly those items with which we associate the highest degree of epistemic certainty. In other words, they are exactly the items with which we have direct experience. It is these items which are also capable of being denoted by logically proper names.[9]

Originally, Russell referred to these items as "sense-data" and characterized them as being "immediately known in sensation." ([22]: 12) Later he referred to them as "percepts," but for current purposes this change in terminology is not significant. Like sense-data, percepts remain both epistemologically privileged and ontologically fundamental.[10]

Underlying Russell's logical atomism there thus appears the clear epistemic goal of constructing less certain items out of more certain ones. From the "primitive data" of experience we are able to construct or infer the existence of physical, mathematical and other objects, and in so doing, our initial uncertainty about these objects is eliminated, or at least greatly reduced.[11]

As a result, we have the following general picture of Russell's theory of knowledge. All propositional knowledge falls into (at least) two general categories, logical knowledge and empirical knowledge. In both cases there exists a foundational class of propositions that serves as a source of premisses for all derived beliefs. These propositions are known directly or intuitively, and include both truths about immediate facts of sensation and truths of logic. In both cases our knowledge of these propositions is, or comes close to being, certain, not only in the sense that these propositions are indubitable, but also in the sense that there exist no other propositions from which we could legitimately infer their falsehood.

These epistemologically privileged propositions are also atomic, both in the sense that they contain no other propositions as parts, and in the sense that the members of any pair of true atomic propositions will be logically independent of one another. Ideally, our logic will perfectly mirror the structure of these (and all other) propositions. Sentences used to express atomic propositions will include both proper names and

[9] It is in this context that Russell's theory of descriptions, which solves the problem of how a non-existent entity can serve as the subject of a proposition, becomes fundamental. His solution is justly famous, not simply because he denies that many denoting phrases function as logically proper names, but because he shows how propositions involving such phrases may be reconstructed to permit thisthis. (See [19])

[10] Initially, Russell distinguished sense-data from sensations, in the sense that sensations were to be mental acts which took sense-data as their objects. However, by the time he wrote *The Analysis of Mind* in 1921 he had given up his belief in the existence of mental acts, largely because he no longer believed such acts to be empirically detectable. As a result, he also abandoned the idea of sense-data, substituting in its place the notion of "percepts."

[11] Of course the project is not without its difficulties. For example, Ayer criticizes the construction of physical objects out of sense-data as follows: "This theory is highly ingenious, but seems to me to fail on the count of circularity. The difficulty is that if the physical object is to be constructed out of its appearances, it cannot itself be used to collect them. The different appearances of the penny, in Russell's example, have first to be associated purely on the basis of their qualities. But since different pennies may look very much alike, and since they may also be perceived against very similar backgrounds, the only way in which we can make sure of associating just those sensibilia that belong to the same penny is by situating them in wider concepts." See [1]: 78.

predicates. Names will have as their meanings the objects they denote. Names may also be distinguished from predicates since, unlike predicates, they are not tied to any particular propositional form or structure. In other words, unlike, say, a two-place predicate which must be accompanied (in any atomic sentence) by exactly two names, names need not be associated with any specific form of predicate. Although human beings may be acquainted with a variety of entities, including universals,[12] the only entities that we may name (as individuals) are the individual sense-data (or percepts) with which we are acquainted. Any individual not known to us by acquaintance may be known only through description.

Knowledge by description also turns out to rely crucially upon knowledge by acquaintance. As Russell summarizes, "I think that all knowledge as to what there is in the world, if it does not directly report facts known through perception or memory, must be inferred from premisses of which, one, at least is known by perception or memory." ([37]: 97–98) In other words, accepting the existence of unperceived objects still requires us to keep sense-data as the ultimate building blocks of the world.[13] The danger is that, in doing so, we may forget that unperceived objects really do exist. As Russell explains,

> I feel that the concept of 'experience' has been very much over-emphasised, especially in the Idealist philosophy, but also in many forms of empiricism. … [Idealists] tend to think that only what is experienced can be known to exist and that it is meaningless to assert that some things exist although we do not know them to exist. … Everybody, in fact, accepts innumerable propositions about things not experienced, but when people begin to philosophise they seem to think it necessary to make themselves artificially stupid. I will admit at once that there are difficulties in explaining how we acquire knowledge that transcends experience, but I think the view that we have no such knowledge is utterly untenable. ([37]: 97)

Russell goes on to observe that each claim involving knowledge by description will have associated with it a degree of risk. This risk arises since each such claim relies upon an inference to the effect that there exists an entity of the type described. This in turn leads us back to Russell's "supreme maxim" that "Whenever possible, logical constructions are to be substituted for inferred entities." ([26]: 155) However, since we dramatically reduce our chance of error whenever we reduce a postulated object to one constructed from objects known by acquaintance, the epistemic pay-off of Russell's

[12] For example, see [22]: 109. Famously, Russell eventually abandoned the idea that the bare self was an object of acquaintance, eliminating it in favour of classes of experiences. For example, compare [22]: 50, 109, with [24]: 73f, and [30]: chs. 5 and 6.

[13] In other words, according to Russell it was a mistake for the Idealists to accept both (1) the claim that sense-data have all the qualities they appear to have, and (2) the claim that all the qualities of sense-data are inevitably manifest. From an epistemological point of view, it is not just certainty concerning the existence of things that is open to question, but also the certainty of claims *about* all such things. (I am grateful to Russell Wahl for this helpful way of putting things.)

program is immediate and clear. It is only by adopting Russell's theory of logical atomism together with his "supreme maxim" that we are able to answer the sceptic.[14]

2. Russell's Method of Analysis

Recalling C. D. Broad's quip that "As we all know, Mr. Russell produces a different system of philosophy every few years," ([3]: 79) many commentators have suggested that Russell flirted with a large variety of unrelated doctrines. For example, such doctrines are regularly cited to include absolute idealism, platonism, logical atomism, phenomenalism, and neutral monism, to name only a few. As a result, the charge is often made that, rather than being a systematic philosopher, Russell was at best a philosophical opportunist, adopting new views and abandoning old ones whenever the need arose.

In contrast to this view, Russell himself saw all of his post-idealist philosophy as having a unity not noticed by many authors.[15] This unity arises as a result of his methodology: philosophical analysis. In fact, Russell often claimed that he had more confidence in his methodology than in any specific philosophical theory or conclusion. As he himself puts it,

> My method invariably is to start from something vague but puzzling, something which seems indubitable but which I cannot express with any precision. I go through a process which is like that of first seeing something with the naked eye and then examining it through a microscope. I find that by fixity of attention divisions and distinctions appear where none at first was visible, just as through a microscope you can see the bacilli in impure water which without the microscope are not discernible. There are many who decry analysis, but it has seemed to be evident, as in the case of the impure water, that analysis gives new knowledge without destroying any of the previously existing knowledge. This applies not only to the structure of physical things, but quite as much to concepts. 'Knowledge,' for example, as commonly used is a very imprecise term covering a number of different things and a number of stages from certainty to slight probability.
>
> It seems to me that philosophical investigation, as far as I have experience of it, starts from that curious and unsatisfactory state of mind in

[14] It is worth noting that this epistemic pay-off effects our knowledge of truths just as it effects our knowledge of things. Once we reduce our knowledge of postulated objects to knowledge of objects known by acquaintance, all that remains is a series of epistemologically secure inferences. As Russell puts it in his [22]: 109 , "[o]ur *derivative* knowledge of truths consists of everything that we can deduce from self-evident truths by the use of self-evident principles of deduction."

[15] This interpretation has also been developed in accounts such as [44], [48], [14], [8], and [10].

> which one feels complete certainty without being able to say what one is certain of. The process that results from prolonged attention is just like that of watching an object approaching through a thick fog: at first it is only a vague darkness, but as it approaches articulations appear and one discovers that it is a man or a woman, or a horse or a cow or what not. It seems to me that those who object to analysis would wish us to be content with the initial dark blur. Belief in the above process is my strongest and most unshakable prejudice as regards the methods of philosophical investigation. ([37]: 98–99)

As Russell sees it, philosophical analysis mirrors the methodology of the natural sciences. It consists of identifying some phenomenon which is not well understood, analyzing this phenomenon in order to develop a theory that explains it, and then testing, altering and refining the theory until it accounts for all relevant data.

In this sense, philosophical analysis is closely related to what Russell calls the "liberal" or "scientific" outlook. As he puts it, "The essence of the liberal outlook lies not in *what* opinions are held, but in *how* they are held: instead of being held dogmatically, they are held tentatively, and with a consciousness that new evidence may at any moment lead to their abandonment. This is the way in which opinions are held in science, as opposed to the way in which they are held in theology." ([35]: 15) Hence the title for the first of Russell's 1914 Lowell lectures, "The Scientific Method in Philosophy."[16]

This allegiance to method mattered a great deal to Russell. For example, here he is in a letter to Lady Ottoline Morrell commenting on a conversation he had with one of his students in 1912:

> I began to talk about how philosophy should be studied—how people ought to have more of the scientific impulse for collecting queer facts, less fear of spending their time on matters not dignified in themselves but important for their consequences, as the man of science does with his test-tubes; how the love of system, since new facts are the enemies of systems, has to be kept rigidly in check, in spite of being a thing every philosopher ought to have; how vital it is to avoid unction and edification and the wish to be literary. Some day I must write on how to study philosophy; I have a lot to say about it. There is so much to be found out by patience and a scientific spirit … [These ideas then] led me on to say how I didn't want to teach a doctrine, but a spirit, an attitude to philosophy. I do care *enormously* about that.[17]

And here he is two years later, continuing to extol the virtues of analysis:

[16] Revised versions of these lectures appeared as chapters in [24]. Russell also used a similar title, "On Scientific Method in Philosophy," several months later when he gave the Herbert Spencer lecture in London. An abridged version of the lecture appears as [27] in [28].

[17] Russell writing to Lady Ottoline Morrell in March of 1912, quoted in [40]: 6.

> Of the prospect of progress in philosophy, it would be rash to speak with confidence. Many of the traditional problems of philosophy, perhaps most of those which have interested a wider circle than that of technical students, do not appear to be soluble by scientific methods. Just as astronomy lost much of its human interest when it ceased to be astrology, so philosophy must lose in attractiveness as it grows less prodigal of promises. But to the large and still growing body of men engaged in the pursuit of science—men who hitherto, not without justification, have turned aside from philosophy with a certain contempt—the new method, successful already in such time-honoured problems as number, infinity, continuity, space and time, should make an appeal which the older methods have wholly failed to make. Physics, with its principle of relativity and its revolutionary investigations into the nature of matter, is feeling the need for that kind of novelty in fundamental hypotheses which scientific philosophy aims at facilitating. The one and only condition, I believe, which is necessary in order to secure for philosophy in the near future an achievement surpassing all that has hitherto been accomplished by philosophers, is the creation of a school of men with scientific training and philosophical interests, unhampered by the traditions of the past, and not misled by the literary methods of those who copy the ancients in all except their merits. ([24]: 242)

Clearly, the very fact that analytic philosophy was as dominant as it was during the 20th century tells us something about Russell's success. Even so, in the current context, there is one specific feature of Russell's method worth emphasizing. This is its hypothetical or inductive character.

Because of the work of scholars such as Peter Hylton[18] and Nicholas Griffin,[19] we now know just how much Russell's philosophy was influenced by his early brush with British idealism. However, like his Cambridge contemporary, G. E. Moore, the young Russell soon broke with his idealist origins. Russell saw that the idealist doctrine of internal relations led to a series of contradictions regarding asymmetrical (and other) relations necessary for mathematics.[20] This led to his defense of realism and to his claim that relations exist independently of the individual natures of their relata. This led, in turn, to his theory of logical atomism, in which all atomic facts exist in isolation from one another. As a result, knowledge is largely to be gathered piecemeal, or by induction, and it is for this reason that all (scientific) knowledge remains uncertain. It is also for this reason that analysis is so useful. Since we obtain knowledge of things independently of the totalities of which they are a part, it is only possible to obtain new knowledge of the whole by first learning something about its constituent parts.

[18] For example, see [12].

[19] For example, see [8].

[20] For example, in his [37]: 42 he comments as follows: "Leibniz gives an extreme example. He says that, if a man living in Europe has a wife in India and the wife dies without his knowing it, the man undergoes an intrinsic change at the moment of her death. This is the kind of doctrine that I was combating."

Knowledge claims obtained through philosophical analysis, like theoretical scientific knowledge in general, are built up "tentatively, and with a consciousness that new evidence may at any moment lead to their abandonment." ([35]: 15)

As Hylton[21] has also pointed out, prior to 1905 Russell writes as if it is the analysis of ideas or concepts that is of primary interest to him.[22] However, with the publication of "On Denoting" Russell shifts from the analysis of concepts to the analysis of propositions[23] and it is from this point that Russell begins to emphasize his claim that the form of a sentence will not, in general, be a helpful guide to the underlying logical form of the proposition expressed by the sentence. Discovering the logical form of a proposition or, as we are more likely to say today, a fact[24] is thus not as straight forward as one originally might have hoped. The discovery of the correct underlying form will be largely a matter of hypothesis and confirmation. In other words, it too will be largely a matter of induction.

Unlike many philosophers who followed him, Russell's theory of analysis is thus intimately related to his realism. On Russell's view, it is real-world facts (or propositions) that are to be studied and analyzed, not mind-independent concepts, or portions of either scientific or natural language.[25] At the same time, philosophical analysis is equally concerned with the discovery of logical form,[26] and it is in this sense that "Philosophy becomes indistinguishable from logic as that word has now come to be used."[27] In other words, it is the connection between real world facts and their underlying logical form that analysis is intended to exhibit.

As we have already seen, analysis, in this sense, is also intimately connected to Russell's epistemology, and to his notion of acquaintance. Analysis is complete only when we discover the elementary, epistemologically privileged atoms known to us through acquaintance. However, as Bill Lycan has pointed out, for the contemporary scientific realist, this connection between ontological and logical atoms is less than straight-forward:

> To a (purely hypothetical) contemporary metaphysician who had never been influenced by Russell, the alleged identity of logical atoms with ontological atoms (and accordingly, that of logical fiction with ontological complexes of some sort) would come as quite a surprise. A philosopher who inclines toward scientific realism thinks of scientific posits such as quarks and gluons as being the basic building blocks of the physical universe at least, but this same philosopher, turning to semantics, would never suppose that the true sentences of English were semantically analyzable

[21] See [13]: 41.

[22] For example, see [17]: 18 and [18]: 111.

[23] See not only [19], but also [33]: ch. 12.

[24] Russell himself equivocates between these two terms.

[25] For example, see both [9] and [13] for helpful surveys of how the term "analysis" was understood by other philosophers, including G. E. Moore, Rudolf Carnap, Ludwig Wittgenstein and W. V. Quine.

[26] For example, see [24]: 42, or [27]: 112.

[27] Cf. [27]: 111–112. See also [31]: 323.

> into logical formulas whose (genuine) referring terms named quarks or gluons … Why, then, did Russell maintain to the contrary that all and only logical atoms are ontological atoms …?
>
> There is a quick and obvious first answer to this question: Russell was not a 'scientific realist' in the sense I have gestured toward, but a phenomenalist. His ontology featured sense-data, not scientific posits such as quarks and gluons. And it is the *epistemological* properties of Russell's sense-data that form the bridge between logical and ontological fictionhood. ([16]: 237–238)

But just how strong is this bridge? Given the large amount of non-deductive detective work required to distinguish between grammatical and logical form, what guarantee do we have that Russell's original goal of achieving epistemic certainty will be satisfied? And given Russell's heavy reliance on induction in justifying both logical and scientific knowledge, how certain can we be about any individual knowledge claim, whether in logic, science, or philosophy?

In his early writings, Russell uses the term "regressive method" to refer to any type of inference that involves the confirmation of hypotheses by their consequences. In the sciences, this type of confirmation is regularly considered to serve as a norm. However, given Russell's overall scientific outlook, this same method is found even in mathematics and philosophy.[28] In the case of mathematics, for example, Russell observes that "Some of the premisses are much less obvious than some of their consequences, and are believed chiefly because of their consequences,"[29] and since "the inferring of premisses from consequences is the essence of induction," it follows that "the method in investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science." ([20]: 274) Yet if so, a high degree of uncertainty is bound to effect even logic and mathematics, and once the regressive method is adopted, Russell's original goal of achieving universal epistemic certainty fades quickly into the background.

Whether he is defending his adoption of the axiom of reducibility in *Principia Mathematica*[30] or his postulates of scientific inference in *Human Knowledge: Its Scope and Limits*[31] over forty years later, Russell's reliance on induction is central to his overall program of work. The result is a unity, not just within Russell's mathematical philosophy, but within his epistemology as well. As Alan Wood memorably puts it, Russell's "work on epistemology was not a kind of subsidiary supplement to his work on mathematical philosophy. It came from the same workshop and was made with the same tools." ([48]: 196)[32] Ultimately, however, this unity would bring with it the downfall of Russell's epistemic foundationalism.

[28] See [14] for a summary of the basic argument.

[29] See [31]: 325. Compare [33]: 16.

[30] See [45]: xiv, 59–60. Compare [37]: 90f.

[31] For a summary, see [36]: ch. 9.

[32] Also see [37]: 12 where we read that it is Russell's goal to "reverse the process which has been common in philosophy since Kant … [namely] to begin with how we know and proceed afterwards to what we know."

3. The Tension between Knowledge and Method

It is well known that, according to Russell, philosophy has two main purposes. The first is to encourage a degree of intellectual modesty; to show that there are things that we thought we knew but do not. As Russell explains, philosophy allows us to see that even our most everyday observations have associated with them questions that are difficult to resolve. As he puts it in *Problems of Philosophy*, philosophy "removes the somewhat arrogant dogmatism of those who have never traveled into the region of liberating doubt, and it keeps alive our sense of wonder by showing familiar things in an unfamiliar aspect." ([22]: 157)

Philosophy's second purpose is to encourage speculation about things that are not yet amenable to scientific investigation. In other words, it is "to keep alive that speculative interest in the universe which is apt to be killed by confining ourselves to definitely ascertainable knowledge." ([22]: 156) Says Russell, "There are a great many things of immense interest about which science, at present at any rate, knows little and I don't want people's imaginations to be limited and enclosed within what can be now known." ([38]: 9f.) As a result, philosophical questions "enlarge our conception of what is possible, enrich our intellectual imagination, and diminish the dogmatic assurance which closes the mind against speculation …." ([22]: 161)

On this view, despite its shared method (in the broad sense) with science, philosophy should be understood as a kind of pre-science, something that occupies us prior to our being able to address matters scientifically or within what, to borrow a term from Kuhn, we might call "normal science." As Russell puts it, "philosophy consists of speculations about matters where exact knowledge is not yet possible. … science is what we know and philosophy is what we don't know." ([38]: 9) It is for this reason that "questions are perpetually passing over from philosophy into science as knowledge advances." ([38]: 9) It is for this same reason that philosophy contains so little that is uncontroversial. As Russell explains, as soon as definite knowledge concerning any subject becomes possible,

> this subject ceases to be called philosophy, and becomes a separate science. The whole study of the heavens, which now belongs to astronomy, was once included in philosophy; Newton's great work was called 'the mathematical principles of natural philosophy.' Similarly, the study of the human mind, which was a part of philosophy, has now been separated from philosophy and has become the science of psychology. Thus, to a great extent, the uncertainty of philosophy is more apparent than real: those questions which are already capable of definite answers are placed in the sciences, while those only to which, at present, no definite answer can be given, remain to form the residue which is called philosophy. ([22]: 155)

According to Russell, "this is a mistake, because knowing how we know is [only] one small department of knowing what we know."

In one sense, then, philosophy and science are on a par: both share a similar method and both have as their goal our understanding of the world. At the same time, philosophy remains pre-scientific in the sense that the questions and issues it addresses are often not yet well-defined, not yet sharp enough to be regularized into normal science. Seen in this light, Gödel's famous comment that "while philosophy analyzes the fundamental concepts, science only uses them"[33] and Einstein's related suggestion that "Science without epistemology is—insofar as it is thinkable at all—primitive and muddled," ([6]: 684) can both be understood as Russellian in temperament.

As Russell tells us, his philosophical methodology therefore emphasizes our ability to move from the vague to the precise, from pre-science to science: "It is a rather curious fact in philosophy that the data which are undeniable to start with are always rather vague and ambiguous. … 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, [and] 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." ([29]: 179–180)

As an example, Russell considers the claim that there are a number of people in this room at this moment. Although in some sense the claim may be undeniable, it turns out that when we try to define what a room is, or what it is for a person to be in a room, or what it is to be a person, we find such tasks remarkably difficult. Moving from the vague to the precise is never easy or straightforward. Says Russell, "If I start with the statement that there are so and so many people in this room, and then set to work to make that statement precise, I shall run a great many risks and it will be extremely likely that any precise statement I make will be something not true at all. So you cannot very easily or simply get from these vague undeniable things to precise things which are going to retain the undeniability of the starting-point." ([29]: 180) Hence we have Russell's famous comment to the effect that "the point of philosophy is to start with something so simple as to not seem worth stating, and to end with something so paradoxical that no one will believe it." ([29]: 193)

For Russell, being a logical atomist is thus really just having a certain kind of method: "It means, in my mind," says Russell, "that the way to get at the nature of any subject matter you're looking at is analysis—and that you can analyze until you get to things that can't be analyzed any further and those would be logical atoms. I call them logical atoms because they're not little bits of matter. They're the ideas, so to speak, out of which a thing is built." ([38]: 12) As we have seen, it is in this sense that Russell claims philosophy to be "indistinguishable from logic." ([27]: 111) It is also in this sense that Russell concludes that methodology, or logic, is fundamental, "and that schools should be characterized rather by their logic than by their metaphysic." ([31]: 323)

As we have seen, Russell is also famous for emphasizing both the fundamental epistemic role that analysis plays, and the similarities between philosophical analysis

[33] Kurt Gödel, paraphrased in [43]: 151.

and scientific method, going so far as to call his method "the scientific method in philosophy."[34] But how are these two aspects of Russell's program to be reconciled? On the one hand, his method of analysis has been designed especially to help answer the sceptic, to reduce epistemologically questionable entities to more certain ones using his "supreme maxim in scientific philosophizing." On the other hand, he also tells us that the very "essence" of his methodology requires that beliefs be held only "tentatively" or provisionally, and that new evidence may at any moment lead to their abandonment. Scientific theories, he reminds us, are accepted only as "useful hypotheses" and that "no sensible person regards them as immutably perfect." ([35]: 18) The result is that, on the one hand, philosophical analysis is designed to secure epistemic *certainty*; but on the other, it has as its consequence universal *uncertainty*.

Of course, one way out of this dilemma might be to conclude that we have misunderstood Russell on either one or both of these points. For example, Russell often states that philosophy is to be distinguished from science, not by its method or its goals, but by its subject matter. Unlike the more subject-specific scientific disciplines, philosophy investigates only the most general aspects of the world, and it is because of this generality that philosophy is able to proceed, as it apparently does, *a priori*. In other words, it may be that our dilemma need not be resolved since, ultimately, it never really arises: theories of epistemic justification, on the one hand, and theories of philosophical method, on the other, would be properly understood as completely distinct theories about completely distinct phenomena. Epistemic justification increases certainty, just as it has been designed to do. Theories *about* epistemic justification, on the other hand, bring with them no such guarantee.

But is this way out of Russell's dilemma successful? If Russell is right, the scientific outlook will permeate *all* branches of rational inquiry, including both science and philosophy. In addition, as Russell is regularly at pains to remind us, "Philosophy, like all other studies, aims primarily at knowledge." ([22]: 154) Unlike Wittgenstein, who in the *Tractatus* tells us that philosophy is distinct from science (cf. [46]: 4.111), Russell sees philosophy as being on a continuum with science. Unlike Wittgenstein's characterization of philosophy as a non-cognitive activity that may be used and practiced, but which ultimately must be "passed over in silence," ([46]: 7) Russell's philosophy, including his theory of method, remains a branch of knowledge proper. Yet if so, it will be sensitive to exactly the same doubts, errors and uncertainties as any other branch of knowledge. It follows that distinguishing between epistemic justification, on the one hand, and theories about epistemic justification, on the other, cannot serve as a way out of our dilemma.

How, then, is this tension between certainty and uncertainty to be resolved?

[34] See note 16 above.

4. Russell's Way out

With the publication, in 1984, of Russell's 1913 *Theory of Knowledge* manuscript, we now know much more about the impact that Wittgenstein's criticisms had on Russell's work in epistemology.[35] These criticisms, it turns out, were just as devastating in epistemology as they had been in logic and mathematics. "I thought of mathematics with reverence," Russell tells us, "and suffered when Wittgenstein led me to regard it as nothing but tautologies."[36] It now appears that the setback was just as great when Russell was faced with Wittgenstein's claim that Russell's theory of judgment was not up to the task of explaining even elementary logical inferences. (See [5]: xxvii) Given Russell's foundationalism, how was logical knowledge to be explained? By June of 1913, Russell would write that Wittgenstein's criticism made him feel "ready for suicide." ([5]: xxvii) Intellectually, it meant that he would be unable to work seriously on his theory of knowledge "for years to come."[37]

In helping us understand why this is so, Ken Blackwell's analysis is particularly telling: "I believe he hoped," Blackwell tells us, that Russell's originally planned book on this topic "would do for epistemology what *Principia* had done for logic and mathematics. In short, while eschewing axioms and symbolic demonstration, he hoped to erect an unimpeachable system, which would function deductively, demonstrating our knowledge of the external world. It was a far grander effort than the book of this title written partly in despair and with drastically pruned ambitions a few months later."[38]

In contrast to this grand hope, Wittgenstein led Russell to the view that his newly formed multiple-relation theory of judgment was an utter failure. (See [7]) Originally introduced by Russell to avoid the problem of false beliefs (i.e., the problem of what it is that our beliefs are about when we believe falsely), the multiple-relation theory ultimately proved to be inconsistent with Russell's theory of types and of logical form. (See [7] and [41]) It seemed that either Russell would have to give up his explanation of how we accept many of even the most elementary judgments, or he would have to give up his resolution of the paradoxes. Since neither option appeared plausible, the result was paralysis.

Yet even more challenging issues were still to come. The main one was that, like logic, philosophy itself was soon to be viewed by Wittgenstein as an activity empty in content. Like logic, epistemology would have to become either an empirical science akin to descriptive psychology or else mere tautology. Thus, just as analytic philosophy itself was about to divide into Quine's naturalized epistemology on the one hand, and Wittgenstein's natural language philosophy on the other, Russell himself

[35] Prior to the release of this material in [23], Russell's only published reference to these criticisms occurred in [39]: vol. 2, 57.

[36] Cf. [34]: 19. See also [37]: 119.

[37] See [2]: 111 and [5]: xxviii.

[38] See [2]: 109. Compare [5]: xxii.

felt forced to choose between these two horns of the dilemma. Initially, at least, the result was intellectually debilitating.

On Wittgenstein's developing view, all meaningful thoughts were truth-functionally analyzable into elementary, atomic propositions. These, in turn, were analyzable into names, predicates and logical forms, all of the appropriate Russellian kind. But how could sentences in philosophy meet these strict requirements? What empirical content does epistemology (or metaphysics, or ethics) have? As Hylton puts it, "if the doctrines of the *Tractatus* are true then they cannot be expressed, so the sentences which (apparently) attempt to express them are nonsensical." ([13]: 45) As Hacker puts it, "The *Tractatus* itself is the swan song of metaphysics, for its propositions are nonsense. There are no philosophical propositions, hence no philosophical knowledge. Philosophy is not a cognitive discipline. Its contribution is not to human knowledge, but to human understanding. The task of philosophy is the activity of logical clarification."[39] And as Urmson so starkly puts it, if Wittgenstein is right, "There are no philosophical propositions." ([42]: 105)

All this was still in the future. In 1913 Russell only knew that his then current theory of knowledge was not up to the task of explaining logical certainty. As he puts it, "My impulse was shattered, like a wave dashed to pieces against a breakwater." ([39]: vol. 2, 57)

Despite this sad conclusion to the 1913 manuscript, our story need not end so unhappily. By the time Russell was able to complete *An Inquiry into Meaning and Truth* in 1940, there was not only change, but resolution. By 1940, Russell had given up the claim that sensations yield direct, error-free knowledge. In other words, he had given up knowledge by acquaintance in the traditional sense and, with it, epistemic foundationalism. In contrast to the robust defense of immediate knowledge given in both *Problems of Philosophy* and *Our Knowledge of the External World*, Russell's new theory of neutral monism opened the door to anti-foundationalism. In effect, this new theory admits that there is no direct perception, and hence that there is no perception free of interpretation, no perception free of potential misinterpretation and potential error. While rejecting coherence in the sense of Neurath and Hempel (cf. [33]: 139–140), Russell comes to accept the view that epistemic justification is based upon "a combination of self-evidence with coherence: sometimes one factor is very much more important than the other, but in theory coherence always plays some part." ([33]: 161) Even at the level of percepts, there is the possibility of error: "When two percepts A and B occur together, the occurrence of a percept closely similar to A on a future occasion may cause an image closely similar to B." ([33]: 161) Thus, as Elizabeth Eames summarizes, by 1940, Russell had

> acknowledged that perceptual experience comes with built-in interpretations due to memory, organic habits, the influence of attention, and the interest of the organism. It seems that it was Russell's reading of behavioristic psychology and a careful attention to experience that brought

[39] Cf. [9]: 19. See, too [46]: 4.112.

> about this change. With this new position the possibility of anchoring descriptive knowledge to its validation in knowledge by acquaintance was excluded. Sensory knowledge became 'the sensory core' of the perceptual experience. ([4]: 123)[40]

Should we be surprised? Perhaps not. Even as early as 1912 Russell was telling us that it is the job of philosophy to

> show us the hierarchy of our instinctive beliefs … It should take care to show that, in the form in which they are finally set forth, our instinctive beliefs do not clash, but form a harmonious system. There can never be any reason for rejecting one instinctive belief except that it clashes with others; thus, if they are found to harmonize, the whole system becomes worthy of acceptance. … by organizing our instinctive beliefs and their consequences, by considering which among them is most possible, if necessary, to modify or abandon, we can arrive, on the basis of accepting as our sole data what we instinctively believe at an orderly systematic organization of our knowledge, in which, though the *possibility* of error remains, its likelihood is diminished by the interrelation of the parts and by the critical scrutiny which has preceded acquiescence. ([22]: 25–26)

Seen in the context of Russell's original foundationalism, such comments are clearly anomalous. At least some instinctive beliefs, after all, were originally intended to be both direct and error-free.[41] But if so, how could such beliefs ever come into conflict with one another? Also, on a foundationalist account, why should some instinctive beliefs be thought more probable than others? And, why would epistemic justification increase once these beliefs are made "to harmonize"?

In contrast, any theory of knowledge that abandons certainty ([33]: 133) and that emphasizes both the importance of coherence and the necessity of integrating scientific and philosophical knowledge ([33]: 131) will have avoided the tension between foundationalism and method. Thus, it is by abandoning his foundationalism that Russell not only resolves the tension between certainty and uncertainty; it is with this single change that he side-steps Wittgenstein's challenge and clears the way for Quine.

Acknowledgement. An abbreviated version of this paper was read at the conference "One Hundred Years of Russell's Paradox" held at the University of Munich in

[40] Even so, in parts of his [33]: 137 ff, Russell remains ambiguous about several of these points. For example, although they are defined as "a subclass of epistemological premisses … which are caused, as immediately *as possible*, by perceptive experiences" (italics added), Russell clearly continues to advocate the acceptance of what he calls "basic propositions" (see especially ch. 10). Later in the book Russell also comments as follows: "There is a tendency—not confined to Neurath and Hempel, but prevalent in much modern philosophy—to forget the arguments of Descartes and Berkeley. It may be that these arguments can be refuted, though, as regards our present question, I do not believe that they can be." ([33]: 143f.)

[41] This is so even though Russell distinguished between instinctive and intuitive knowledge. For example, see [22]: ch. 11. Again, I am grateful to Russell Wahl for reminding me of this point.

June, 2001. I would like to extend my thanks to Godehard Link and the other members of the conference's organizing committee for their support. I would also like to thank Melinda Hogan, Peter Hylton, Alan Richardson and Russell Wahl for their helpful comments and suggestions.

References

[1] Ayer, Alfred J.: 1972. Bertrand Russell as a philosopher. *Proceedings of the British Academy* 58: 127–151. Repr. in [15], vol. 1, 65–85.

[2] Blackwell, Kenneth: 1981. The early Wittgenstein and the middle Russell. In: Irving Block (ed.), *Perspectives on the Philosophy of Wittgenstein*, Cambridge, MA: MIT Press, 1–30. Repr. in [15] vol. 1, 96–123.

[3] Broad, Charles D.: 1924. Critical and speculative philosophy. In: John H. Muirhead (ed.), *Contemporary British Philosophy*, London: George Allen & Unwin, 75–100.

[4] Eames, Elizabeth R.: 1972. Russell on 'What there is'. *Revue Internationale de Philosophie* 26: 483–498. Repr. in [15] vol. 3, 115–127.

[5] Eames, Elizabeth R.: 1984. Introduction. In: [23], xv-xlv.

[6] Einstein, Albert: 1949. Reply to criticisms. In: Paul A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, London: Cambridge University Press, 663–688.

[7] Griffin, Nicholas: 1986. Wittgenstein's criticism of Russell's theory of judgement. *Russell* 4: 132–145.

[8] Griffin, Nicholas: 1991. *Russell's Idealist Apprenticeship*. Oxford: Clarendon.

[9] Hacker, Peter M. S.: 1998. Analytic philosophy: What, whence, and whither? In: Anat Biletzki and Anat Matar (eds.), *The Story of Analytic Philosophy*. London: Routledge, 3–34.

[10] Hager, Paul J.: 1994. *Continuity and Change in the Development of Russell's Philosophy*. Dordrecht: Kluwer.

[11] Hamilton, Edith, and Huntington Cairns (eds.): 1961. *The Collected Dialogues of Plato*. Princeton: Princeton University Press.

[12] Hylton, Peter: 1990. *Russell, Idealism, and the Emergence of Analytic Philosophy*. Oxford: Clarendon.

[13] Hylton, Peter: 1998. Analysis in analytic philosophy. In: Anat Biletzki and Anat Matar (eds.), *The Story of Analytic Philosophy*, London: Routledge, 38–55.

[14] Irvine, Andrew D.: 1989. Epistemic logicism and Russell's regressive method. *Philosophical Studies* 55: 303–327. Repr. in [15], vol. 2, 172–195.

[15] Irvine, Andrew D.: 1999. *Bertrand Russell: Critical Assessments*, 4 vols. London: Routledge.

[16] Lycan, William: 1981. Logical atomism and ontological atoms. *Synthese* 46: 207–229. Repr. in [15] vol. 3, 237–238.

[17] Russell, Bertrand: 1900. *A Critical Exposition of the Philosophy of Leibniz*, 2nd edn. London: George Allen & Unwin, 1937.

[18] Russell, Bertrand: 1903. *Principles of Mathematics*, 2nd edn. London: George Allen & Unwin, 1937.

[19] Russell, Bertrand: 1905. On denoting. In: *Logic and Knowledge*, London: George Allen & Unwin, 1956, 41–56.

[20] Russell, Bertrand: 1907. The regressive method of discovering the premises of mathematics. In: *Essays in Analysis*, London: George Allen & Unwin, 1973, 272–283.

[21] Russell, Bertrand: 1911. Knowledge by acquaintance and knowledge by description. In: [28], 209–232.

[22] Russell, Bertrand: 1912. *Problems of Philosophy*. New York and Oxford: Oxford University Press, 1997.

[23] Russell, Bertrand: 1913. *Theory of Knowledge: The 1913 Manuscript*. Collected Papers of Bertrand Russell, vol. 7. London: George Allen & Unwin, 1984.

[24] Russell, Bertrand: 1914. *Our Knowledge of the External World*. Chicago and London: Open Court.

[25] Russell, Bertrand: 1914. On the nature of acquaintance. In: *Logic and Knowledge*, London: George Allen & Unwin, 1956, 127–174.

[26] Russell, Bertrand: 1914. The relation of sense-data to physics. In [28]: 145–179.

[27] Russell, Bertrand: 1914. On scientific method in philosophy. In [28]: 97–124.

[28] Russell, Bertrand: 1918. *Mysticism and Logic and Other Essays*. London: George Allen & Unwin.

[29] Russell, Bertrand: 1918. The philosophy of logical atomism. In: *Logic and Knowledge*, London: George Allen & Unwin, 1956, 177–281.

[30] Russell, Bertrand: 1921. *The Analysis of Mind*. London: George Allen & Unwin.

[31] Russell, Bertrand: 1924. Logical atomism. In: *Logic and Knowledge*. London: George Allen & Unwin, 1956, 323–343.

[32] Russell, Bertrand: 1927. *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner.

[33] Russell, Bertrand: 1940. *An Inquiry into Meaning and Truth*. London: George Allen & Unwin.

[34] Russell, Bertrand: 1944. My mental development. In: Paul A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, 3rd edn, New York: Tudor, 1951, 1–20.

[35] Russell, Bertrand: 1947. Philosophy and politics. In: *Unpopular Essays*, New York: Simon and Schuster, 1950, 1–20.

[36] Russell, Bertrand: 1948. *Human Knowledge: Its Scope and Limits*. London: George Allen & Unwin.

[37] Russell, Bertrand: 1959. *My Philosophical Development*. London: Routledge, 1993.

[38] Russell, Bertrand: 1960. *Bertrand Russell Speaks His Mind*. New York: Bard Books.

[39] Russell, Bertrand: 1967, 1968, 1969. *The Autobiography of Bertrand Russell*, 3 vols. London: George Allen & Unwin.

[40] Slater, John: 1988. Russell's conception of philosophy. *Russell* n.s. 8: 163–178. Repr. in [15]: vol. 3, 1–20.

[41] Sommerville, Stephen: 1980. Wittgenstein to Russell (July, 1913): 'I am very sorry to hear … my objection paralyses you'. In: Rudolf Haller and Wolfgang Grassl (eds.), *Language, Logic and Philosophy: Proceedings of the Fourth International Wittgenstein Symposium*, Vienna: Hölder-Pichler-Tempsky, 182–188.

[42] Urmson, James O.: 1956. *Philosophical Analysis*. London: Oxford University Press.

[43] Wang, Hao: 1987. *Reflections on Kurt Gödel*. Cambridge, MA: MIT Press.

[44] Weitz, Morris: 1944. Analysis and the unity of Russell's philosophy. In: Paul A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, 3rd edn, New York: Tudor, 1951, 55–121.

[45] Whitehead, Alfred North, and Bertrand Russell: 1910. *Principia Mathematica to *56*. Cambridge: Cambridge University Press, 1962.

[46] Wittgenstein, Ludwig: 1921. *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul, 1961.

[47] Wittgenstein, Ludwig: 1995. *Cambridge Letters*. Oxford: Blackwell.

[48] Wood, Allan: 1959. Russell's philosophy: A study of its development. In [37]: 189–205.

Department of Philosophy
University of British Columbia
1866 Main Mall E370
Vancouver BC
Canada V6T 1Z1

E-mail: a.irvine@ubc.ca

Paradoxes in Göttingen

Volker Peckhaus

Abstract. There are two periods in the early discussion of the paradoxes in Göttingen. The first period up to 1903 saw the struggle with Georg Cantor's unintended sets (Cantor's Paradox) in the course of which David Hilbert formulated a paradox of his own. The logical significance of this paradox, but also of Ernst Zermelo's Paradox, similar to, but independent of Bertrand Russell's Paradox, was not recognized. So they remained unpublished. When Russell's Paradox was published by Russell himself and by Gottlob Frege, Hilbert had to change his axiomatic programme significantly. Now the paradoxes were also discussed in the circle around the Göttingen philosopher Leonard Nelson. A fruit of these discussions was the semantical paradox found by Kurt Grelling.

1. Introduction

In 1903 the mathematical world was struck by Russell's Paradox twice over. Bertrand Russell himself published it under the heading "The Contradiction" in chapter 10 of his *Principles of Mathematics* [40]. Almost at the same time Gottlob Frege (1848–1925) referred to Russell's Paradox in the postscript of the second and final volume of his *Grundgesetze der Arithmetik* [7], admitting that the foundational system of the *Grundgesetze* had turned out to be inconsistent.

Frege sent a copy of this volume to his Göttingen colleague David Hilbert (1862–1943), and received an astonishing reply. Hilbert told Frege that "this example," as he called the paradox, had already been known before in Göttingen. In a footnote he added, "I believe Dr. Zermelo discovered it three or four years ago after I had communicated my examples to him," and continued

> I found other even more convincing contradictions as long as four or five years ago; they led me to the conviction that traditional logic is inadequate and that the theory of concept formation needs to be sharpened and refined.[1]

[1] See [9]: 51. German original in [8]: 79–80.

In view of this remarkable claim of Hilbert's a number of questions can be raised:

- What exactly were these paradoxes said to have been known in Göttingen already before the publication of Russell's Paradox?
- What was Ernst Zermelo's (1871–1953) role in this story?
- Why did Hilbert and Zermelo refrain from publishing the paradoxes they had found?
- What was the impact of Russell's Paradox, if any, in Göttingen?
- In particular, who discussed the paradox in Göttingen? Did such discussions go beyond the bounds of mathematical discourse?

Most of these questions have already been given quite satisfactory answers. They show that there is just as little ground for the "standard" story in the history of set theory—according to which Russell's paradox led to the axiomatization of set theory—as for assuming that in Göttingen at least there was no impact at all, a conclusion which could well be drawn from Hilbert's self-confident statement in his letter to Frege. However, the impact was not like a sudden blow that immediately divided the mathematical history into pre and post paradox mathematics; rather, the paradoxes had their effect only gradually, after some time of thinking them over, but then sparked new and important insights into the nature of mathematics.

This paper will not present new historical material on the impact of Russell's Paradox in Göttingen. It aims at surveying and combining the results of previous research on the matter, thereby illustrating the complexity of the story of the paradoxes.

The paper will be divided into two parts. In the first part the period up to 1903 will be considered, especially Hilbert's exchange with Cantor on unintended sets, and the paradoxes of Hilbert and Zermelo. The second part is devoted to the effects caused by the publication of Russell's Paradox. It will cover the period between 1903 and 1908. In this part the discussion of Russell's Paradox among Göttingen philosophers standing close to the Hilbert circle will also be presented.

2. The Paradoxes up to 1903

2.1. Cantor's Paradox

Hilbert's first contact with the so-called "paradoxes of set theory" can be traced back to his correspondence with Georg Cantor (1845–1918) between 1897 and 1900.[2] Already in the first of his letters, dated 26 September 1897 ([6]: no. 156, 388–389),

[2]For a comprehensive discussion of this correspondence [36]: 147–166. Extracts are published in [6].

Cantor proved that the totality of alephs does not exist, i. e., that this totality is not a well-defined, finished set [*fertige Menge*]. If it were taken to be a finished set, a certain larger aleph would follow on this totality. So this new aleph would at the same time belong to the totality of all alephs, and not belong to it, because of being larger than all alephs (*ibid.*: 388).

Hilbert and Cantor were thus concerned with what later became known under the name "Cantor's Paradox". But in fact, Cantor was not really dealing with paradoxes and their resolution, but with non-existence proofs using *reductio-ad-absurdum* arguments.[3] He disproved the existence of the totality of all cardinals by showing that the assumption of its existence contradicts his definition of a set as the comprehension of certain well-distinguished objects of our intuition or our thinking into a whole ([4], *cf.* [5]: 282). The totality of all cardinals (and of all ordinals) cannot be thought of as *one* such thing, in contrast to actual infinite objects like transfinite sets.

Hilbert's letters to Cantor have not been preserved, but we can derive Hilbert's opinion from published talks. These talks show that Hilbert took the doubts about the existence of the set of all cardinals seriously, obviously not convinced by Cantor's non-existence proof. Hilbert saw an alternative to this proof in a suitable axiomatization of set theory. In his 1900 paper "On the Concept of Number" [15], Hilbert's first paper on the foundations of arithmetic, he gave a set of axioms for arithmetic, and claimed that only a suitable modification of known methods of inference would be needed for proving the consistency of the axioms. If this proof were successful, the existence of the totality of real numbers would be shown at the same time. In this context he referred to Cantor's problem of whether the system of real numbers is a consistent, or finished, set. He stressed:

> Under the conception above, the doubts which have been raised against the existence of the totality of all real numbers (and against the existence of infinite sets in general) lose all justification; for by the set of real numbers we do not have to imagine the totality of all possible laws according to which the elements of a fundamental sequence can proceed, but rather—as just described—a system of things whose internal relations are given by a *finite and closed* set of axioms […], and about which new statements are valid only if one can derive them from the axioms by means of a finite number of logical inferences.[4]

He also claimed that the existence of the totality of all powers or of all Cantorian alephs could be disproved, i. e., in Cantor's terminology, that the system of all powers is an inconsistent (not finished) set (*ibid.*).

Hilbert took up this topic again in his famous 1900 Paris lecture on "Mathematical Problems."[5] In the context of his commentary on the second problem concerning the consistency of the axioms for arithmetic he used the same examples from Cantorian set

[3] We follow in this evaluation [25] and [10].

[4] See [21]: 1095. German original [15]: 184.

[5] See [14], English translations in [16] and [20].

theory and the continuum problem as in the earlier lecture. "If contradictory attributes be assigned to a concept," he wrote, "I say, that mathematically the concept does not exist" ([20]: 1105).

According to Hilbert a suitable axiomatization would avoid the contradictions resulting from the attempt to comprehend absolute infinite multiplicities as units, because only those concepts had to be accepted which could be deduced from an axiomatic basis.

2.2. Hilbert's Paradox

Hilbert's remarks at these prominent places show that he was worried by the problems causing Cantor to distinguish between finished or consistent sets and multiplicities that are no sets. Unrestricted comprehension, a natural way of forming sets, had been removed by a stipulation made necessary by the fact that a contradiction would arise if that were not done. A suitable axiomatization, however, would exclude these problematic multiplicities from the outset. Hilbert was worried not simply by the fact that these contradictions occurred in set theory, but by the pragmatic aspect that methods used by mathematicians as a matter of course had proven to lead to unintended, contradictory results. And this pragmatic aspect furthermore shows that not only Cantorian set theory could be implicated, but also other domains of mathematics. Hilbert gave evidence to this by formulating a paradox of his own which was "purely mathematical" in the sense that it did not make use of notions from Cantor's theory of cardinals and ordinals. This paradox was known in Göttingen as "Hilbert's Paradox," the one he obviously referred to in his letter to Frege, and which Otto Blumenthal mentioned in his biographical note in the third volume of Hilbert's *Collected Works* [1].[6]

Hilbert never published the paradox. He discussed it, however, in a lecture course on the "Logical Principles of Mathematical Thinking" he gave in Göttingen in the summer semester of 1905. It is preserved in two sets of notes [18], [19]. Part B of these notes, on "The Logical Foundations", starts with a comprehensive discussion of the paradoxes of set theory. It begins with metaphorical considerations of the general development of science:

> It was, indeed, usual practice in the historical development of science that we began cultivating a discipline without many scruples, pressing onwards as far as possible, that we thereby, however, then ran into difficulties (often only after a long time) that forced us to turn back and reflect on the foundations of the discipline. The house of knowledge is not erected like a dwelling where the foundation is first well laid-out before the erection of the living quarters begins. Science prefers to obtain comfortable rooms as quickly as possible in which it can rule, and

[6] For the history of Hilbert's Paradox and its reconstruction, *cf.* [29] and [35] (with an edition and English translation of Hilbert's presentation of the paradox).

> only subsequently, when it becomes clear that, here and there, the loosely joined foundations are unable to support the completion of the rooms, science proceeds in propping up and securing them. This is no shortcoming but rather a correct and healthy development.[7]

Although contradictions are quite common in science, Hilbert continued, in the case of set theory they seem to be different, because there they have a tendency to bend towards theoretical philosophy. In set theory the common Aristotelian logic and its standard methods of concept formation had been used without hesitation. And these standard tools of purely logical operations, especially the subsumption of concepts under a general concept, now proved to be responsible for the new contradictions.

Hilbert elucidated these considerations by presenting three examples, the Liar Paradox, "Zermelo's Paradox," as the Russell–Zermelo Paradox was called in Göttingen at that time which will be discussed below, and the paradox of his own, which was, according to Hilbert, of purely mathematical nature ([18]: 210). Hilbert expressed his opinion that this paradox

> appears to be especially important; when I found it, I thought in the beginning that it causes invincible problems for set theory that would finally lead to the latter's eventual failure; now I firmly believe, however, that everything essential can be kept after a revision of the foundations, as always in science up to now. I have not published this contradiction, but it is known to set theorists, especially to G. Cantor.[8]

The paradox is based on a special notion of set which Hilbert introduces by means of two set formation principles starting from the natural numbers. The first principle is the *addition principle* (*Additionsprinzip*). In analogy to the finite case, Hilbert argued that the principle can be used for uniting two sets together "into a new conceptual unit", a new set that contains the elements from both sets. This operation can be extended: "In the same way, we are able to unite several sets and even infinitely many into a union." The second principle is called the *mapping principle* (*Belegungsprinzip*). Given a set $\mathcal{M}$, he introduces the set $\mathcal{M}^{\mathcal{M}}$ of *self-mappings* (*Selbstbelegungen*) of $\mathcal{M}$ to itself. A self-mapping is just a total function ("transformation") which maps the elements of $\mathcal{M}$ to elements of $\mathcal{M}$.[9]

Now, he considers all sets which result from the natural numbers "by applying the operations of addition and self-mapping an arbitrary number of times." By the addition principle, which allows us to build the union of arbitrary sets, one can "unite them all into a sum set $\mathcal{U}$, which is well-defined." In the next step the mapping principle is applied to $\mathcal{U}$, and we get $\mathcal{F} = \mathcal{U}^{\mathcal{U}}$ as the set of all self-mappings of $\mathcal{U}$. Since $\mathcal{F}$ was built from the natural numbers by using the two principles alone, Hilbert concludes that it has to be contained in $\mathcal{U}$. From this fact he derives a contradiction.

[7] See [19]: 122, published in [29]: 51.

[8] See [18]: 204 and [35]: 164.

[9] In standard set theory, $\mathcal{M}^{\mathcal{M}}$ is isomorphic to $2^{\mathcal{M}}$, the set of all functions from $\mathcal{M}$ to $\{0, 1\}$, which, in turn, is isomorphic to $\mathcal{P}(\mathcal{M})$, the power set of $\mathcal{M}$.

Since "there are 'not more' elements" in $\mathcal{F}$ than in $\mathcal{U}$ there is an assignment of the elements u_i of $\mathcal{U}$ to elements f_i of $\mathcal{F}$ such that all elements of f_i are used. Now one can define a self-mapping g of $\mathcal{U}$ which differs from all f_i. Thus, g is not contained in $\mathcal{F}$. Since $\mathcal{F}$ was assumed to contain all self-mappings, we have a contradiction. In order to define g, Hilbert used Cantor's diagonalization method. If f_i is a mapping u_i to $f_i(u_i) = u_{f_i^{(i)}}$ he chooses an element $u_{g^{(i)}}$ different from $u_{f_i^{(i)}}$ as the image of u_i under g. Thus, we have $g(u_i) = u_{g^{(i)}} \neq u_{f_i^{(i)}}$ and g "is distinct from any mapping f_k of $\mathcal{F}$ in at least one assignment."[10]

Hilbert finishes his argument with the following observation:

> We could also formulate this contradiction so that, according to the last consideration, the set $\mathcal{U}^{\mathcal{U}}$ is always bigger [of greater cardinality][11] than $\mathcal{U}$ but, according to the former, is an element of $\mathcal{U}$.

Hilbert's Paradox is closely related to Cantor's Paradox. Both Cantor and Hilbert construct "sets" which lead to contradictions. This is shown with the help of Cantor's diagonalization argument. However, the ways in which these "sets" are constructed differ essentially. According to Cantor ([3]: § 11; *cf.* [5]: 195–197), there are three principles for the generation of cardinals. The first principle ("erstes Erzeugungsprinzip") concerns the generation of real whole numbers [*reale ganze Zahlen*, i. e., ordinal numbers] by adding a unit to a given, already generated number. The second principle allows the formation of a new number, if a certain succession of whole numbers with no greatest number is given. This new number is imagined as the limit of this succession. Cantor adds a third principle, the inhibition or restriction principle ("Hemmungs- oder Beschränkungsprinzip") which guarantees that the second number class has not only a higher cardinality than the first number class, but exactly the next higher cardinality. Considering Cantor's general definition of a set, cited above, one can justly ask whether the set of all cardinals, the set of all ordinals and the universal set of all sets, are sets according to this definition, i. e., whether an unrestricted comprehension is possible. Cantor denies this, justifying his opinion with the help of a *reductio ad absurdum* argument, but he doesn't exclude the possibility of forming the paradoxes by provisions in his formalism.

Hilbert, on the other hand, introduces two alternative set formation principles, the addition principle and the mapping principle, but they lead to paradoxes as well. In avoiding concepts from transfinite arithmetic Hilbert believes that the purely mathematical nature of his paradox is guaranteed. For him, this paradox appears to be much more serious for mathematics than Cantor's, because it concerns an operation that is part of everyday practice of working mathematicians.

---

[10] Hilbert's notation $u_{g^{(i)}}$ is somewhat clumsy. In fact, it is enough to say that $g(u_i) = v_i$ for an element v_i of $\mathcal{U}$ with $v_i \neq f_i(u_i)$.

[11] Remark later added in Hilbert's hand in Hellinger's lecture notes.

2.3. Zermelo's Paradox

Ernst Zermelo came to Göttingen in 1897 in order to work on his *Habilitation*. His special fields of competence were the calculus of variations and mathematical physics, such as thermodynamics and hydrodynamics.[12] Under the influence of Hilbert he changed the focus of his interests to set theory and foundations. He became Hilbert's collaborator in the foundations of mathematics. His first set-theoretical publication on the addition of transfinite cardinals dates from 1901 [44], but as early as the winter semester 1900/1901 he gave a lecture course on set theory in Göttingen. It is possible that he may have found the paradox while preparing this course. He referred to it in the famous polemical paper "A New Proof of the Possibility of a Well-Ordering" of 1908 [45]. There Zermelo noted that he had found the paradox independently of Russell, and that he had mentioned it to Hilbert and other people already before 1903. And indeed, among the papers of Edmund Husserl (1859–1938), professor of philosophy in Göttingen until 1916, a note in Husserl's hand was found, partially written in Gabelsberger shorthand, saying that Zermelo had informed him on 16 April 1902 that the assumption of a set M that contains all of its subsets m, m', ... as elements, is an inconsistent set, i. e., a set which, if treated as a set at all, leads to contradictions.[13] Zermelo's message was a comment on a review that Husserl had written on the first volume of Ernst Schröder's (1841–1902) *Vorlesungen über die Algebra der Logik* [41]. Schröder had criticized George Boole's interpretation of the symbol 1 as the class of everything that can be a subject of discourse (the universe of discourse, universal class).[14] Husserl had dismissed Schröder's argumentation as sophistic ([23]: 272), and was now advised by Zermelo that Schröder was right concerning the matter, but not in his proof.

In his own recollections, communicated to Heinrich Scholz in 1936, Zermelo saw the origins of his paradox in discussions in the Hilbert circle. At that time Heinrich Scholz was working on the papers of Gottlob Frege which he had acquired for his department at the University of Münster. He had found Hilbert's letter to Frege, mentioned above, and now asked Zermelo what paradoxes Hilbert referred to in this letter.[15] Zermelo answered that the set-theoretic paradoxes were often discussed in the Hilbert circle around 1900, and he himself had at that time given a precise formulation of the paradox which was later named after Russell.[16]

As mentioned above, Hilbert discussed it in his 1905 lecture course, presenting it as a "purely logical" example—probably more convincing for non-mathematicians—

[12] On Zermelo's activities in Göttingen *cf.* esp. [26], [28], [29]: 76–122.

[13] Critical edition in ([24]: 399). English translation in [38].

[14] See [41]: 245. Schröder referred to Boole's definition of the universe of discourse and his interpretation of the symbol 1, *cf.* [2]: 42–43.

[15] Heinrich Scholz to Zermelo, dated Münster, 5 April 1936, University Archive Freiburg i. Br., Zermelo papers, C 129/106.

[16] Zermelo to Scholz, dated Freiburg i. Br., 10 April 1936, Institut für mathematische Logik und Grundlagenforschung, Münster, Scholz papers.

whereas his own (purely mathematical) example seamed to be more decisive for the mathematician ([18]: 210).

3. Reflecting on the Paradoxes, 1903–1908

3.1. The Mathematicians' Reaction to Russell's Paradox

Given the naïve attitude towards the contradictions of set theory, the publication of Russell's Paradox could not easily be ignored because the paradox unveiled the consequences of this special kind of contradictions for general logic. Before 1903 Hilbert suggested avoiding or circumventing the contradictions with the help of an axiomatic reformulation of set theory. This suggestion was kept after 1903, but he had to extend it considerably. His axiomatic programme was deeply involved since its centrepiece, the consistency proof for the arithmetical axioms by means of standard mathematical and logical methods, required logic itself to be free of contradictions.

In his seminal "Grundlagen der Geometrie" [13] Hilbert had proved the consistency of Euclidean geometry under the provision of the consistency of arithmetic. The consistency of arithmetic, however, had still to be shown. In his talk "Über den Zahlbegriff" [15] Hilbert claimed that the consistency proof for arithmetic only required a suitable modification of known methods of inference ([15]: 184), an opinion which soon proved to be overoptimistic. In the Paris problems talk, he included the consistency proof for arithmetic as the second among the problems discussed [14]. Frege's disappointed admission now made it obvious that this consistency proof could not be done using means just proved to be inconsistent.

The talks on foundations of mathematics and logic at meetings of the Göttingen Mathematical Society give evidence for the new activities released by the publication of the paradoxes.[17] In 1901 and 1902 the main speaker was Hilbert himself, giving papers on special problems in the axiomatic foundation of geometry. Furthermore, Edmund Husserl gave his "double lecture" on completeness and definiteness of axiomatic systems (26 November and 10 December 1901). In 1903 Ernst Zermelo discussed Frege's concept of number as presented in the second volume of the *Grundgesetze der Arithmetik* (12 May 1903). Obviously the Göttingen mathematicians felt a need to go deeper into Frege's failed theory. Hilbert spoke about the axiomatic standpoint in the foundations of arithmetic emphasizing the principle of contradiction as the "pièce de résistance" (27 December 1903). Early in 1904 H. Fleischer familiarized the Göttingen mathematicians with Italian positions in foundations, reporting, e. g., on Giuseppe Peano's (1858–1932) *Arithmetices principia, nova methododo exposita*

[17] According to the reports in the *Jahresbericht der Deutschen Mathematiker-Vereinigung* **11** (1901)–**14** (1905).

(Peano *1889*) (19 January and 23 February 1904). Several lectures on the axiomatization of set theory were given, especially by E. Zermelo in 1904, 1906 and 1908. And on 7 February 1905 William Henry Young (1863–1942) reported on Russell's *Principles of Mathematics* and Russell's Paradox.

These activities are indications for a deep revision of Hilbert's axiomatic programme. The programme before 1903 consisted in axiomatizing certain fields of mathematics in which foundations had been questioned, reducing their consistency to the consistency of arithmetic. Set theory was not an integral methodological part of this programme, but was among the fields of mathematics waiting to be axiomatized. After 1903 it became clear that the axiomatization of arithmetic required an axiomatization of logic and set theory. One precondition was to gain competence in modern logic, i. e., both, the Fregean mathematical logic and G. Boole's and E. Schröder's algebra of logic. Therefore logic, formerly regarded as a basic philosophical subdiscipline, came into the focus of the Göttingen mathematicians. On Hilbert's initiative, Zermelo was awarded with a stipendiary lectureship on "Mathematical Logic and Related Fields" in 1907. Zermelo delivered the first German lecture course on mathematical logic which was based on a ministerial commission, in the summer semester of 1908, *cf.* [28, 29, 30, 32].

Now logic and set theory were integrated into the axiomatic programme, i. e., securing the logical and set-theoretical grounds was seen as a necessary precondition for proving the consistency of arithmetic. Arithmetic was treated in a rather logicist manner, i. e., as soon as the consistency of logic and set theory (the last dealing with the infinite in mathematics) were shown, there would only be one step more to prove the consistency of arithmetic. This was expressed by Hilbert in his rather obscure demand of "a partially simultaneous development of the laws of logic and arithmetic" called for in his Heidelberg talk on the foundations of logic and arithmetic, Hilbert's first published reaction to the paradoxes ([17]: 170). These ideas were considerably deepened in the lecture course on the logical principles of mathematical thinking mentioned above.

Hilbert's original approach to reducing the consistency of any set of mathematical axioms to the consistency of arithmetic was thus replaced by a three step programme to create consistent sets of axioms for logic, set theory, and then arithmetic. One of the first fruits of this revised programme was Ernst Zermelo's formulation of the set-theoretic axioms published in 1908. Zermelo, however, had to confess that he was not yet able to prove "the 'consistency,' no doubt very essential, of my axioms" ([46]: 262).

3.2. The Philosophical Discussion in the Nelson Circle

One of the participants in Hilbert's summer lecture course of 1905 was the 23 year-old philosopher Leonard Nelson (1882–1927), who had received his doctorate in

philosophy in Göttingen in July 1904.[18] He was already head of a philosophical school, the *Neue Fries'sche Schule* that was devoted to critical philosophy in the spirit of Jacob Friedrich Fries (1773–1843). Nelson was still a student when he founded the new series of the *Abhandlungen der Fries'schen Schule* as a forum for his circle, assisted by his older friend Gerhard Hessenberg (1873–1925), then lecturer for mathematics at the Academy for Military Technology in Charlottenburg near Berlin, and by the physiologist Karl Kaiser (1861– c. 1933). Hessenberg was a known geometer, but he also gained recognition as a set theorist after having published the first textbook on set theory [12].

In June 1905 Nelson sent a letter to Hessenberg commenting on Hilbert's lecture "Über die Grundlagen der Arithmetik" [17], and he expressed his disappointment about Hilbert's ideas. Rather perplexed he wrote: "In order to remove the contradictions in set theory, he [i.e., Hilbert] intends to reform (not set theory but) logic. Well, we shall see, how he will do it."[19] Hessenberg answered quite to the point:[20]

> I do not at all consider it as paradoxical that one has to reform logic in order to make set theory free of contradictions. First of all it is not yet possible to separate logic sharply from arithmetical considerations. Secondly, however: If there are paradoxes in set theory, then either the inferences are not correct or the concepts generated are contradictory.

In both cases, Hessenberg continued, it is a logical task to uncover the mistakes. According to the laws of logic, a thing a either does or does not fall under the concept b. No other principle is needed for the concept of a set. Hessenberg stressed that Hilbert very much strengthened the requirements for building concepts in order to avoid the resulting paradoxes.

In Hilbert's lecture course Nelson learned more about the paradoxes, not about Russell's, but about its Göttingen variant, Zermelo's Paradox, and about Hilbert's Paradox. It is instructive to observe the growing significance of Russell's name in the correspondence between Hessenberg and Nelson. In June 1905 Hessenberg, who was at that time going to write his textbook on set theory, recalled that Nelson had spoken about "Hilbert's Paradox of the set of sets that belong to themselves," obviously misconceiving what Nelson had told him.[21] In the same month Nelson directed Hessenberg's interest towards Russell's *Principles of Mathematics*, a book in which the foundations of all mathematical disciplines and especially of set theory were discussed,

[18] For more details on Nelson and his circle, *cf.* [29]: 123–154, with further hints on literature. On the philosophical discussion of the paradoxes in Göttingen, especially the emergence of Grelling's Paradox, see [34].

[19] Nelson to Hessenberg, dated Göttingen, 16 June 1905, Bundesarchiv, Abt. Potsdam, Nelson Papers, 90 Ne 1, fol. 50–53.

[20] Hessenberg to Nelson, dated Grunewald, 26 June 1905, Archiv der sozialen Demokratie, Nelson Papers, 1/LN AA 000270.

[21] Hessenberg to Nelson, dated Grunewald, 26 June 1905, Archiv der sozialen Demokratie, Nelson Papers, *ibid*.

as Nelson wrote.[22] In February 1906 Nelson sent Russell's book to Hessenberg whose judgement on its set-theoretical parts was, however, scathing. He had found almost nothing that made any impression on him. He regarded Russell's logicistic standpoint (contrary to that of Dedekind) as utterly ridiculous, and criticized that Russell expatiated "the completely vague contradiction of the set of all sets that do not contain themselves in a whole chapter, but found only a few unimportant words on the considerable contradiction of the set of all ordinals."[23]

It is astonishing to see that, although the paradoxes became a widely discussed topic, in the beginning no one made efforts to study the relevant books of Russell or Frege in greater detail, in Nelson's circle at least not before late 1905 or early 1906.

In contrast to Hessenberg, Russell's Paradox did electrify the more philosophically minded members of Nelson's circle. In June 1906 Heinrich Goesch (1880–1930, *cf.* [29]: 137–140), wrote the following letter to his friend Leonard Nelson (mentioning other members of Nelson's circle):[24]

> Grelling informed me of Russel[l]'s Paradox in logic concerning the concept impredicable. I succeeded in solving the same, and then I learned from Berkowski of another paradox also from Russel[l] in set theory, concerning the concept of a set not belonging to itself. For this paradox my resolution holds as well. Therefore Berkowski, who told me that the mathematicians working in set theory have no resolution for the paradox up to now, thought that the matter might be not unimportant. I therefore would like to write a short paper, and I would like to ask you to tell me in which book of Russel[l]'s these paradoxes can be found and to which German presentations [*Ausformungen*] one should refer. I think that the matter will be finished in a few days.

Of course, Goesch's paper was not written in a few days. A first version of the announced manuscript was not completed before spring 1907.[25] Nelson was sceptical and commissioned his closest collaborator, the mathematics student Kurt Grelling (1886–1943)[26] to check Goesch's ideas.

In this period of discussion, several other members of the circle were involved in attempts to solve the paradoxes. One of them was Otto Meyerhof (1884–1951),[27] the 1923 Nobel laureate for medicine and physiology; another was Alexander Rüstow (1885–1963),[28] after World War II one of the fathers of the German social economy.

[22] Nelson to Hessenberg, dated Grunewald, 7 February 1906, Archiv der sozialen Demokratie, Nelson Papers, 1/LN AA 000271.

[23] Hessenberg to Nelson, dated Grunewald, 7 February 1906, Bundesarchiv, Abt. Potsdam, 90 Ne 1, no. 389, fol. 54*f*.

[24] Goesch to Nelson, undated (Munich, 14 June 1906), Archiv der sozialen Demokratie, Nelson Papers, 1/LN AA 000255.

[25] In the correspondence with Nelson (Archiv der sozialen Demokratie, Nelson Papers, LN AA000255) a postal receipt of delivery for a manuscript can be found, dated 23 April 1907.

[26] On Grelling's tragical biography, *cf.* [31, 33].

[27] *Cf.* [29]: 135–137.

[28] *Cf.* [29]: 140–142.

Rüstow asked Nelson whether he could publish his doctoral thesis in the *Abhandlungen der Fries'schen Schule*.[29] It had the characteristic title *Der Lügner. Theorie, Geschichte und Auflösung des Russellschen Paradoxons* (*The Liar. Theory, History, and Resolution of Russell's Paradox*). Nelson refused, because he rejected Rüstow's resolution.[30] Rüstow's thesis was published not before 1910, and it is also characteristic that in the title of the published version the reference to Russell's Paradox was omitted [39].

It was during Grelling's and Nelson's struggle with Goesch's "resolution" that Grelling and Nelson compiled material for a joint paper which was finally published in 1908 [11]. They looked for the basic logical conditions for the occurrences of the paradoxes, and distinguished between

> the task of a proper "resolution" of the paradox, i.e., the task of unveiling the underlying appearance, and the task of a "correction," i.e., the task of avoiding the paradox by introducing new, consistent concepts. Such a correction cannot be considered to be a resolution, because the paradoxical objects, if they exist at all, are not eliminated by stopping work on them. ([11]: 314)

Most importantly, Grelling discovered new paradoxes, among them the semantic "heterological" paradox, today known under Grelling's name. It runs in its original version as follows:

> Let $\varphi(M)$ be the word that denotes the concept defining M. This word is either an element of M or not. In the first case we will call it "autological" in the other "heterological."[31] Now the word "*heterological*" is itself either autological or heterological. Suppose it to be autological; then it is an element of the set defined by the concept that is denoted by itself, hence it is heterological, contrary to the supposition. Suppose, however, that it is heterological; then it is not element of the set defined by the concept that is denoted by itself, hence it is not heterological, again against the supposition. ([11]: 307)

It is this paradox of Grelling's that Frank Plumpton Ramsey (1903–1930) in [37] wrongly attributed to Hermann Weyl (1885–1955), who had mentioned it in *Das Kontinuum* as "a well-known paradox, essentially coming from Russell" and had discussed it as "scholasticism of the worst kind" ([43]: 2). History has shown that Weyl's judgement did not do justice to the importance of Grelling's Paradox.

Acknowledgement. I would like to thank Marcello Ghin (Paderborn) and Godehard Link (Munich) for helpful comments on a previous version, and Daniel Mook (Munich), for his suggestions concerning my English. This paper very much profited from

[29] *Cf.* Nelson to Hessenberg, dated Westend, 23 January 1908, Bundesarchiv, Abt. Potsdam, 90 Ne 1, no. 389, fol. 155–157.

[30] Nelson to Rüstow, dated Göttingen, 9 February 1908, Bundesarchiv Koblenz, NL 169, Rüstow.

[31] "Short", e.g., is autological; "long", heterological; "English" is autological; "German", heterological.

an earlier collaboration with Reinhard Kahle on "Hilbert's Paradox", the result of which has been published as [35]. All translations from the German are mine, unless otherwise stated.

References

[1] Blumenthal, Otto: 1935. Lebensgeschichte. In: D. Hilbert, *Gesammelte Abhandlungen*, vol. III. Berlin: Springer, 388–429. Second edition, Berlin: Springer, 1970.

[2] Boole, George: 1854. *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton & Maberly. Reprinted Dover, New York: c. 1958.

[3] Cantor, Georg: 1883. Ueber unendliche, lineare Punktmannichfaltigkeiten [5th pt.]. *Mathematische Annalen* 21: 545–591. Reprinted in [5]: 165–209.

[4] Cantor, Georg: 1895 & 1897. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen* 46/49: 207–246. Reprinted in [5]: 282–351.

[5] Cantor, Georg: 1932. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Edited by E. Zermelo. Berlin: Springer. Reprinted Hildesheim: Olms, 1962 and Berlin: Springer, 1980.

[6] Cantor, Georg: 1991. *Briefe*. Edited by H. Meschkowski/W. Nilson. Berlin: Springer.

[7] Frege, Gottlob: 1903. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. vol. 2. Jena: Hermann Pohle. Reprinted together with vol. 1, Hildesheim: Olms, 1966.

[8] Frege, Gottlob: 1976. *Wissenschaftlicher Briefwechsel*. Edited by G. Gabriel *et al*. Hamburg: Felix Meiner.

[9] Frege, Gottlob: 1980. *Philosophical and Mathematical Correspondence*. Edited by G. Gabriel *et al*. Oxford: Basil Blackwell.

[10] Garciadiego Dantan, Alejandro R.: 1992. *Bertrand Russell and the Origins of the Set-theoretic "Paradoxes"* . Basel: Birkhäuser.

[11] Grelling, Kurt and Leonard Nelson: 1908. Bemerkungen zu den Paradoxieen von Russell und Burali-Forti. *Abhandlungen der Fries'schen Schule* n. s. 2, no. 3: 301–334.

[12] Hessenberg, Gerhard: 1906. Grundbegriffe der Mengenlehre. Zweiter Bericht über das Unendliche in der Mathematik. *Abhandlungen der Fries'schen Schule* n. s. 1, no. 4: 479–706.

[13] Hilbert, David: 1899. Grundlagen der Geometrie. In: *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen*, Leipzig: Teubner, 1–92. Recent edition [22].

[14] Hilbert, David: 1900. Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900. *Nachrichten von der königl. Gesellschaft der Wissenschaften zu Göttingen*. Mathematisch-physikalische Klasse aus dem Jahre 1900, 253–297. English translation [16]. Extract in [20].

[15] Hilbert, David: 1900. Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8: 180–184. English translation in [21].

[16] Hilbert, David: 1902. Mathematical Problems. *Lecture Delivered before the International Congress of Mathematicians at Paris in 1900*. Translated by M. W. Newson. *Bulletin of the American Mathematical Society* 8: 437–479.

[17] Hilbert, David: 1905. Über die Grundlagen der Logik und der Arithmetik. In: A. Krazer (ed.), *Verhandlungen des Dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174–185.

[18] Hilbert, David: 1905. *Logische Principien des mathematischen Denkens*. Lecture course in the summer semester 1905, lecture notes by Ernst Hellinger. Library of the Mathematics Seminar of the University of Göttingen.

[19] Hilbert, David: 1905. *Logische Principien des mathematischen Denkens*. Lecture course in the summer semester 1905, lecture notes by Max Born. Niedersächsische Staats- und Universitätsbibliothek Göttingen, Cod. Ms. D. Hilbert 558a.

[20] Hilbert, David: 1996. Mathematical Problems. In: W. Ewald (ed.), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. 2. Oxford: Clarendon Press, 1096–1105.

[21] Hilbert, David: 1996. On the Concept of Number. In: W. Ewald (ed.), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. 2. Oxford: Clarendon Press, 1089–1095.

[22] Hilbert, David: 1999. *Grundlagen der Geometrie. Mit Supplementen von Paul Bernays*. 14th ed. by M. Toepell. Stuttgart & Leipzig: Teubner.

[23] Husserl, Edmund: 1891. Review of [41]. *Göttingische gelehrte Anzeigen*, 243–278. Critical edition in [24]: 3–43.

[24] Husserl, Edmund: 1979. *Aufsätze und Rezensionen (1890–1910), mit ergänzenden Texten*. Edited by B. Rang. Husserliana vol. XXII, The Hague: Nijhoff.

[25] Moore, Gregory H. and Alejandro Garciadiego: 1981. Burali-Forti's Paradox: A Reappraisal of its Origins. *Historia Mathematica* 8: 319–350.

[26] Moore, Gregory H.: 1982. *Zermelo's Axiom of Choice. Its Origins, Development and Influence*. Studies in the History of Mathematics and Physical Sciences 8. Berlin: Springer.

[27] Peano, Giuseppe: 1889. *Arithmetices principia, nova methododo exposita*. Augustae Taurinorum: Bocca.

[28] Peckhaus, Volker: 1990. 'Ich habe mich wohl gehütet, alle Patronen auf einmal zu verschießen.' Ernst Zermelo in Göttingen. *History and Philosophy of Logic* 11: 19–58.

[29] Peckhaus, Volker: 1990. *Hilbertprogramm und Kritische Philosophie. Das Göttinger Modell interdisziplinärer Zusammenarbeit zwischen Mathematik und Philosophie*. Studien zur Wissenschafts-, Sozial- und Bildungsgeschichte der Mathematik 7. Göttingen: Vandenhoeck & Ruprecht.

[30] Peckhaus, Volker: 1992. Hilbert, Zermelo und die Institutionalisierung der mathematischen Logik in Deutschland. *Berichte zur Wissenschaftsgeschichte* 15: 27–38.

[31] Peckhaus, Volker: 1993. Kurt Grelling und der Logische Empirismus. In: R. Haller and F. Stadler (eds.), *Wien – Berlin – Prag. Der Aufstieg der wissenschaftlichen Philosophie. Zentenarien Rudolf Carnap – Hans Reichenbach – Edgar Zilsel*. Veröffentlichungen des Instituts Wiener Kreis 2, Wien: Hölder-Pichler-Tempsky, 362–385.

[32] Peckhaus, Volker: 1994. Logic in Transition: The Logical Calculi of Hilbert (1905) and Zermelo (1908). In: D. Prawitz and D. Westerståhl (eds.), *Logic and Philosophy of Science in Uppsala. Papers from the 9th International Congress of Logic, Methodology and Philosophy of Science*. Synthese Library 236. Dordrecht: Kluwer, 311–323.

[33] Peckhaus, Volker: 1994. Von Nelson zu Reichenbach. Kurt Grelling in Göttingen und Berlin. In: L. Danneberg *et al.* (eds.), *Hans Reichenbach und die Berliner Gruppe*, Braunschweig & Wiesbaden: Vieweg & Sohn, 53–73.

[34] Peckhaus, Volker: 1995. The Genesis of Grelling's Paradox. In: I. Max and W. Stelzner (eds.), *Logik und Mathematik. Frege-Kolloquium Jena 1993*. Perspectives in Analytical Philosophy 5, Berlin: de Gruyter, 269–280.

[35] Peckhaus, Volker and Reinhard Kahle: 2002. Hilbert's Paradox. *Historia Mathematica* 29: 157–175.

[36] Purkert, Walter and Hans J. Ilgauds: 1985. *Georg Cantor*. Biographien hervorragender Naturwissenschaftler, Techniker und Mediziner 79. Leipzig: Teubner.

[37] Ramsey, Frank P.: 1926. The Foundations of Mathematics. *Proceedings of the London Mathematical Society* (2) 25: 338–384.

[38] Rang, Bernhard and W. Thomas: 1981. Zermelo's Discovery of the 'Russell Paradox'. *Historia Mathematica* 8: 15–22.

[39] Rüstow, Alexander: 1910. *Der Lügner. Theorie, Geschichte und Auflösung*. Ph. D. thesis Erlangen 1908. Leipzig: Teubner.

[40] Russell, Bertrand: 1903. *The Principles of Mathematics*, Cambridge: The University Press. London: Allen & Unwin, ^{2}1937. Paperback edition London: Routledge, 1992.

[41] Schröder, Ernst: 1890. *Vorlesungen über die Algebra der Logik (exakte Logik)*. vol. 1. Leipzig: Teubner. Reprinted as vol. 1 of the second edition Bronx, NY: Chelsea, 1966.

[42] van Heijenoort, Jean: 1967. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press.

[43] Weyl, Hermann: 1918. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Leipzig: Veit & Comp.

[44] Zermelo, Ernst: 1901 Ueber die Addition transfiniter Cardinalzahlen. *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse aus dem Jahre 1901*, 34–38.

[45] Zermelo, Ernst: 1908. Neuer Beweis für die Möglichkeit einer Wohlordnung. *Mathematische Annalen* 65: 107–128. Translated in [42]: 183–198.

[46] Zermelo, Ernst: 1908. Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen* 65: 261–281. Translated in [42]: 199–215.

Universität Paderborn
Fakultät für Kulturwissenschaften - Philosophie
Warburger Str. 100
33098 Paderborn
Germany

E-mail: peckhaus@hrz.upb.de

David Hilbert and Paul du Bois-Reymond: Limits and Ideals

David Charles McCarty

Abstract. Paul du Bois-Reymond (1831–1889), younger brother of physiologist Emil du Bois-Reymond, was among the most accomplished and influential mathematicians in the second half of the 19th Century. He made substantial contributions to the areas (we would today call) differential equations, calculus of variations, functional analysis, topology and nonstandard analysis, not to mention the foundations of mathematics. Paul du Bois-Reymond may have been the first to employ an explicitly diagonal argument and is credited with introducing into analysis the notions of dense set and choice sequence. In his "General Function Theory" (1882), du Bois-Reymond presented and defended a skeptical philosophy of mathematics and argued informally, on the basis of that philosophy, that mathematics contains absolutely undecidable statements. In the first decades of the 20th Century, mathematician David Hilbert explicitly attacked Emil du Bois-Reymond's famous "Ignorabimus" in several articles and lectures. It is here argued that Paul du Bois-Reymond's ideas, although not so famous, afforded a natural target for the positions Hilbert took up in his proof-theoretic Program, perhaps a more natural target than Brouwer's intuitionism.

1. Hilbert's Program and Brouwer's Intuitionism

Doubtless, David Hilbert and his fellow proof theorists devised the mature Hilbert Program of the 1920s and 30s to form, in large part, a defensive bulwark against Brouwer's intuitionism. Even so, the Program was neither born nor is it well understood exclusively as a countermeasure to intuitionism. It was born of and nurtured by what one could reasonably call "Hilbert's Project," his decades-long engagement with and advocacy for the true epistemology, proper pedagogy and cultural significance of the mathematical sciences. The means and ends of that Project informed Hilbert's lectures and writings, as is plain from his 1919–20 lectures, *Nature and Mathematical Knowledge* [*Natur und mathematisches Erkennen*] wherein we find the following announcement.

> [I]t is among the noblest tasks of philosophy to investigate the questions how knowledge comes to be, what it is and what it takes as its goal. I will handle my material in the light of these questions. My lecture should,

> therefore, form a kind of preparation for an epistemology [*Erkenntnistheorie*]. ([16]: 3)

Early and late, intellectual ingredients to the Program were garnered from the Project. One such ingredient was Hilbert's Axiom of Solvability: every well-posed mathematical problem admits a resolution: either a definite 'Yes,' a definite 'No,' or a convincing proof that no solution will be forthcoming. Hilbert believed that his proof theory would vindicate the axiom, but his statement of it antedated the proof theory of the mature Program. It featured prominently in the Problems Address of 1900, delivered seven years before Brouwer completed his PhD dissertation and eight years before Brouwer was to launch his first assault on the validity of logical laws. At that time, prior to the First World War, the Program was still in short trousers; and neither Program nor Project could have been a reply to Brouwer, Weyl and company, for the intuitionists then counted as no large threat to mathematical peace in Göttingen.

Hilbert's stated intentions for his postwar Program notwithstanding, one cannot always make good sense of his ideas as reasonable anti-intuitionistic countermeasures, as contributory to *prima facie* effective arguments against Brouwer's views. For one thing, Brouwer could not have accepted Hilbert's leading premises. Metatheorems that various formal systems representing higher mathematics are consistent and complete, demonstrations of which were objects of Hilbert's so strenuous labors, were to be obtained via strictly finitistic reasoning. This required *inter alia* that those metatheorems and their proofs be couched in a language carrying a particular interpretation that Hilbert took some pains to elucidate in his writings. On that interpretation, certain sentences formalized with unbounded quantifiers were to count as incomplete communications: either a manner of mathematical shorthand for other finitistically complete communications or procedures for listing finitistically complete communications. According to Hilbert, such general statements, as read by the finitist, cannot be negated significantly.

> For example, the statement that if a is a numerical symbol, then $a + 1 = 1 + a$ is universally true, is from our finitary perspective *incapable of negation*. ([19]: 144)

Hilbert also insisted that, in a language so interpreted,"We cannot write down number-signs or introduce abbreviations for infinitely many numbers" ([17]: 1123). Further, finitistically admissible denoting terms were to refer exclusively to the perceived shapes of intuited finite sequences: "The objects of [finitistic] number theory are for me—in direct contrast to Dedekind and Frege—the signs themselves, whose shapes can be generally and certainly recognized by us" ([17]: 1121). Brouwer could not in consistency have embraced these finitistic restrictions. This had less to do with the logic at work in the conjectured proofs than with the meanings of the statements that would comprise them. Unbounded universal quantifications in intuitionistic arithmetic were perfectly complete communications; in general, Brouwer thought them to assert the existence of abstract mathematical operations with infinite domains and

ranges. Although they would not, in general, provide confirmatory instances of the *tertium non datur*, they would admit of meaningful negation. Moreover, there was no prohibition in intuitionism against writing down "number-signs or abbreviations for infinitely many numbers" so long as the infinities in question were subject to intuitionistic treatment. Third, according to Brouwer, "The first act of intuitionism [is] the complete separation of mathematics from mathematical language and, accordingly, from the linguistic phenomena of theoretical logic. Intuitionistic mathematics is a languageless construction carried out by the human mind" ([4]: 21). Brouwer dubbed the mathematics licensed by the first act "separable mathematics," that is, mathematics pursued in total separation from language. Therefore, his intuitionistic mathematics, at least in its separable aspect, could not rely for its meaning on forms or shapes of physically realizable sign sequences. Lastly, in Brouwer's mathematical universe, there was no place for the concrete signs whose forms afforded ultimate objects to Hilbert's metamathematics, because the intuitionistic universe of pure mathematics was a universe entirely of mental constructions, and hence abstract to that extent.

Hilbert was well aware that the majority of statements made by everyday mathematicians working in, say, analysis lie beyond the pale of finitistic rephrasing, requiring for their formalizations unbounded universal and existential quantifiers thought to range over uncountable domains. On the foundational design set out in such papers and addresses as *On the Infinite* [19], these statements were to count neither as contentfully finitistic nor as denotationally meaningful. Such meaning as they would bear was to be all inference and no reference or, more accurately, no reference not exhausted by inference. Drawing an analogy between infinitary statements that extend contentful mathematics and ideal points and lines that projective geometers add to extend the Euclidean plane, Hilbert applied the epithet 'ideal' to infinitary statements, and incorporated them into his map of the mathematical world as formulae adjoined to strictly finitistic theories so as to improve their deductive efficacy.

By contrast, a crucial pressure point for the intuitionistic critique of conventional mathematics via a critique of inference would vanish on the assumption that infinitary statements such as the classical Bolzano–Weierstrass Theorem (that every infinite, bounded set of real numbers has an accumulation point) which Hilbert would have treated as ideal, carry no referential contents beyond their inferential roles. For Brouwer, one major difficulty haunting statements of that sort was that the claims they make about real numbers cannot be backed up by reliable forms of inference, forms of inference the cogency of which are responsible to the referential meanings of those claims. Terms in significant mathematical claims, both elementary and higher-order, are to refer to mental constructions existing prior to inferences and to which inferences are wholly responsible. The failure of traditional inference, with its dependence on the *tertium non datur* and other nonconstructive logical rules, lies in its inability to track the course of mental construction. An application of *tertium non datur* to a statement giving a suitable construction may well convert it into a statement to which no construction can correspond. This critique of inference opens the door wide to intuitionism's most characteristic and most daring feature: in its higher-order reaches,

intuitionistic mathematics contradicts, not just formally but also contentually, hallowed results of classical mathematics. Intuitionists of Brouwerian stripe understood and still understand the proposition "every total function from the real numbers into the real numbers is continuous" to be a theorem of intuitionistic real analysis contradicting, and striking at the root of, classical analysis. It denies, e.g., the existence of step functions defined over all real numbers. These functions, too, fail to exist because they cannot correspond to appropriate features of the underlying constructive reality.

Third, on the conception of Hilbertian proof theory most familiar to contemporary scholars, Hilbert assumed that proofs would be finitary objects of the same general sort as intuited strings or sequences. *A fortiori*, every proof would have to terminate in a finite number of steps. To quote him, "In our present investigation, proof itself is something concrete and displayable" ([17]: 1127). Hilbert's hope was to prove consistency for formal theories by reducing to evident absurdity the assumption that the theories in question contained finite proofs of contradictions, proofs that could be fully laid out for examination. Brouwer would have had little of this. In his efforts [1] to prove the correctness of Bar Induction (induction on certain well-founded trees) Brouwer proposed an analysis of proofs on which intuitionistically acceptable proofs are in general infinitistic and, hence, impossible to display and survey exhaustively. In consequence, any metatheorem Hilbert proved showing that no finite proof of a contradiction exists would have been insufficient, assuming Brouwer able to countenance the meaning Hilbert assigned to it, to convince Brouwer that an informal mathematical theory contains no hidden contradiction. For the latter would have allowed that a contradiction in a mathematical discipline might only come to light as the conclusion of an infinite proof.

Next, Hilbertian proof-theoretic efforts were to be applied, not to theories in a mathematical *Umgangssprache*, but to their formalized counterparts in artificial languages. Even an intuitionistically allowable proof that a formal theory representing results of traditional mathematics is minimally coherent might neither encourage nor discomfit a proper Brouwerian. Early intuitionists held that mathematical language, artificial or otherwise, was no appropriate guide to mathematical thought. Such an intuitionist might have offered (but did not, so far as I know) the following analogy. We can allow that a particular painting of an event is, as a painting, visually coherent. We can see it, take it in and understand it. From this it does not follow that what the painting depicts ever took place, or, if it did take place, did so in the manner portrayed. In these respects, relations between an historical event, say Wolfe's death at Quebec, and a painting of it parallel those between mathematical thought and mathematical language. Just as Wolfe's demise occurred before West's canvas ever stood on his easel, intuitionistic mathematical thought comes first and mathematical language later, if at all. Like the historical painter, the speaker of a mathematical language can only tag along afterwards to put things on the lexical canvas, perhaps inaccurately, for posterity. Worse, all painting requires a measure of distortion, the truncation of a limitless, multidimensional reality for the sake of the limited, two-dimensional picture plane. Brouwerians believed mathematical language to harbor analogous distortions.

Expressions of such logical laws as the *tertium non datur* are at best distorted images of their mental originals; they record little or nothing real in constructive thought. Hence, a consistency proof for a formal theory that includes the formalized *tertium non datur* would not underwrite the intuitionistic cogency of the thoughts, if any, the theory captures. In general, language is to play the role neither of conductor nor of guide in what really matters: the unfolding of mathematical thought, an unfolding that is a direct grasping of mathematical objects themselves rather than a vain effort to knit their shadows, painted in words and projected on a narrow screen of language.

Worth special mention is an issue on which Hilbertians and Brouwerians would have agreed, an issue that much exercised Hilbert: the autonomy of mathematics. With the Program and the circumambient Project, Hilbert and his followers hoped to insure that mathematics would solve (and be acknowledged to have solved) by strictly mathematical means all serious foundational problems pertaining to it. The truth of its basic claims as well as the cogency and efficacy of its methods of proof were to be so guaranteed that pressing metaphysical and epistemological worries about mathematics were properly banished by an application of mathematics itself. No other discipline, whether metaphysics or psychology, was to get a last foundational word in, as far as the validity of pure mathematics was concerned. On this issue of mathematical autonomy, the intuitionists stood in seeming agreement with Hilbert, as is apparent from [3]. From the intuitionistic viewpoint, answers to all basic foundational questions regarding mathematics looked to be attainable via mathematical insight, by an exercise of which we were to learn that all mathematical objects are mental constructions and all mathematical proofs result from constructive acts.

Certainly, Hilbert devised the mature Program, in part, to thwart a feared Brouwerian *Putsch*. That Hilbert's dialectical countermeasures failed to meet the intuitionistic assault head-on puts Hilbert, in this respect, in no very bad company; some of his brightest contemporaries struggled to grasp Brouwer's revolutionary ideas and did not always succeed. That said, it is no vain exercise in counterfactual history to adumbrate a position on the foundations of mathematics to which Hilbert's Program stood in more perfect opposition, a position welcoming to Hilbertian premises but unfriendly to their much-sought-after conclusions.

That position would ask (1) that mathematical thought be a manipulation of representations at least partially rule-governed and not, as the intuitionists held, an engagement with mathematical objects or constructions themselves; (2) that proofs be finite in length; (3) that the field of mathematical intellection be divisible into infinitary and finitary zones, the latter maintaining an epistemic priority over the former; (4) that mathematical claims refusing finitistic reconstrual not denote in ways unreflected in their inferential powers, but be treated as limits or ideals adjoined to the finitistic realm; (5) that the Axiom of Solvability for well-posed mathematical problems fail; (6) that mathematics not be autonomous but survive only under the foundational aegis of some nonmathematical science; and (7) that modern higher mathematics, taken as a whole, be inconsistent.

2. Paul du Bois-Reymond: Mathematician and Philosopher

At the close of the 19th Century, Paul du Bois-Reymond was counted among the most successful and influential of European mathematicians. In the index of Ernest William Hobson's classic textbook, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series* [20], no mathematician received more page references than du Bois-Reymond. The index listed Cantor, Dedekind, Dini, Hardy, Lebesque and Weierstrass; only Cantor approached du Bois-Reymond in number of citations. Histories credit du Bois-Reymond with introducing the terms 'extremum' and 'integral equation' into mathematics. He used the word 'metamathematics' in his 1890 monograph *On the Foundations of Knowledge in the Exact Sciences* [*Über die Grundlagen der Erkenntnis in den exakten Wissenschaften*] [12] in roughly its contemporary meaning. He devised a notation for rates of growth of real-valued functions that is a direct ancestor to the "big O" notation of contemporary computer science. Under the title 'limits of indefiniteness,' he defined the *lim inf* and *lim sup* of infinite series and proved basic results governing them. (Cauchy had introduced these notions but du Bois-Reymond may have been the first to recognize their full import.) A major theorem bearing his name states that, if a trigonometric series is convergent and the function defined is Riemann integrable, then the coefficients of the series are precisely those of Fourier. In 1868, he formulated and proved the Second Mean Value Theorem for definite integrals, a result to which his name is now attached but which Dini once ascribed to Weierstrass. In 1873, he constructed a continuous function with divergent Fourier series at every point of a dense set, thus refuting conjectures of Dirichlet and Riemann. In 1875, he described dense sets under the title 'pantachisch,' from the Greek for 'everywhere.' He later claimed, against Georg Cantor, priority in their discovery. (In his textbook, Hobson awarded the laurels to du Bois-Reymond.) Cantor presented his first diagonal proof to the public in 1891 in *On an elementary question of set theory* [*Über eine elementare Frage der Mannigfaltigkeitslehre*] [6]; Paul du Bois-Reymond had been there well before him, having published a plainly diagonal argument in an 1875 article on approximation by infinitesimals [9]. His greatest invention was probably the *Infinitärcalcül* or infinitary calculus, an original, non-Cantorean account of infinite and infinitesimal sizes as first-class entities representing the rates of growth of real-valued functions rather than the magnitudes of collections.

David Paul Gustav du Bois-Reymond was born in Berlin on 2 December 1831. He began his academic career in medicine and physiology at Zürich; his elder brother Emil was a famous physiologist. While in Zürich, Paul collaborated on an important study of the blindspot of the eye. Later, he turned to mathematical physics and pure mathematics, tackling problems of partial differential equations. Du Bois-Reymond came to hold professorial appointments at Heidelberg, Freiburg, Berlin and Tübingen, where he was successor to Hermann Hankel. Hilbert was certainly familiar with du Bois-Reymond's mathematical research; we know that the young Hilbert visited him at least once in Berlin. Hilbert was also in touch with du Bois-Reymond's unique philosophy of mathematics, for du Bois-Reymond was widely recognized as a leading

critic of efforts to arithmetize analysis, as Alfred Pringsheim's article in the *Encyclopedia of Mathematical Sciences* [*Encyklopädie der mathematischen Wissenschaften*] [22] confirms. Du Bois-Reymond died in Freiburg on 7 April 1889, having succumbed to kidney disease while on a train trip.

Emil and Paul du Bois-Reymond played major roles in the *Ignorabimusstreit*, a spirited public debate over skepticism in the natural sciences. Emil's 1872 address to the Organization of German Scientists and Physicians, *On the limits of our knowledge of nature* [*Über die Grenzen des Naturerkennens*] [8], both sparked the debate and baptized it. The address closed with the dramatic pronouncement, "In the face of the puzzle over the nature of matter and force and how they should be conceived, the scientist must, once and for all, resign himself to the far more difficult, renunciatory doctrine, '*Ignorabimus*' [we shall never know]" ([8]: 130). Emil argued that natural science is inherently incomplete in that there are pressing foundational questions concerning important phenomena to which science will never find adequate answers. The address unleashed a whirlwind of argument and counterargument in the press and learned journals that lasted well into the 20th Century. Rudolf Carnap, Ernst Mach, Moritz Schlick and Ludwig Wittgenstein joined the ranks of du Bois-Reymond's critics ([14]: 99–102). As late as 1930, Richard von Mises attacked Emil du Bois-Reymond's lecture on behalf of logical positivism, denouncing Emil's skeptical tropes as *Scheinprobleme* ([27]: 119). Edith Stein employed the word '*Ignorabimus*' in her summer lectures of 1932 as a general term for psychological questions that empirical science would be unable to answer ([25]: 167). Emil du Bois-Reymond's agnosticism was the original *Ignorabimus* against which Hilbert so often railed. Hilbert's denunciation of it loomed large in the Problems Address ([5]: 7) as well as in his final public statement, the Königsberg talk of 1930 ([18]: 378–387). The latter concluded with a direct reference to Emil's 1872 lecture: "In general, unsolvable problems don't exist. Instead of the ridiculous *Ignorabimus*, our motto is, by contrast, 'We must know. We will know"' ([18]: 387). Those optimistic lines, "We must know. We will know," are inscribed on Hilbert's burial monument in Göttingen.

Paul du Bois-Reymond's 1882 monograph *General Function Theory* [*Die allgemeine Functionentheorie*] [11] and the posthumously published *On the Foundations of Knowledge in the Exact Sciences* [12] were in part devoted to sowing skepticism in the garden of pure mathematics. In those works, he expounded a philosophy of mathematics both highly original and remarkably prescient, no mere transliteration into mathematical terms of his brother's agnosticism about physical science. Paul du Bois-Reymond showed himself a forceful critic of arithmetization and logicism and, in that respect as well as others, a direct ancestor of Brouwer. Indeed, the extent to which his ideas anticipated those of Brouwer is one, but hardly the sole, index of the value of du Bois-Reymond's philosophy. In *General Function Theory*, Paul drew a clear distinction between actually and potentially infinite sets and, recognizing that the existence of potential but nonactual infinities makes demands on logic, called into question the general validity of the *tertium non datur*. Also, he may have been the first to describe in print lawless sequences, Cauchy sequences the successive terms of

which cannot be generated by any predetermined rule or procedure, and to attempt to demonstrate their existence. To elucidate the idea of lawlessness, he imagined sequences whose terms are given by throws of a die: "One can also think of the following means of generation for an infinite and lawless number. Every place [in the sequence] is determined by a throw of the die. Since the assumption can surely be made that throws of the die occur throughout eternity, a conception of lawless number is thereby produced" ([11]: 91). Doctrines of lawless sequences still comprise a chapter in the intuitionistic theory of the continuum; contemporary intuitionists have recourse to the very same die-casting analogy to explain their approach to the subject ([26]: 645).

Du Bois-Reymond devised arguments that are close kin to Brouwer's 'weak' counterexamples. He believed that information about the physical world could be so encoded in sequences that, if such an encoding sequence were governed by a law, a knowledge of that law would yield up predictions about the universe that would otherwise be impossible to make. Were we aware of laws for the development of those sequences, he reasoned, we would be able to answer correctly questions about the precise disposition of matter at any point in space and at any time in the past. He wrote, "If we think of matter as infinite, then a constant like the temperature of space is dependent on effects that cannot be cut off at any decimal place. Were its sequence of terms to proceed by a law of formation, then this law would contain the history and picture of all eternity and the infinity of space" ([11]: 91–92). He concluded that, since we shall never possess comprehensive physical knowledge, we will never have laws for the sequence giving the precise temperature of space. The similarity to Brouwer's 'weak' counterexamples should be plain: the intuitionist argues that we will never have a law for sorting real numbers into two categories, "equal to zero" and "unequal to zero," on pain of harvesting, from that law, fruit from the intuitionistic tree of good and evil, that is, knowledge sufficient to prove the *tertium non datur*. By Brouwer's lights, this is knowledge we cannot possess.

A good part of *General Function Theory* was written in the form of a dialogue between two imaginary mathematicians, Idealist and Empiricist. The Idealist championed a conception of the geometrical continuum on which its basic constituents can be transcendent and will include both infinite and infinitesimal magnitudes. The Empiricist restricted consideration to those points, line segments and relations among them that are immanent and available to intuition. Du Bois-Reymond believed that our current and future best efforts at the philosophies of mathematics and of mind will discern only these two distinct, mutually inconsistent, fundamental outlooks on the foundations of mathematics, and no final decision between them will ever be reached. A knockdown mathematical argument will not be devised for favoring one over the other. Now or later, the choice between them is largely a matter of scientific temperament. According to du Bois-Reymond, mathematics, which is the scientific study of magnitude, can pass its ultimate foundational test only by being carried back to the touchstone of the continuum and these two conceptions of it. Consequently, he thought the literary device of debate between Idealist and Empiricist to reflect a natural divide and an eternal dispute within human mathematical cognition.

Du Bois-Reymond reasoned that this permanent intellectual dualism engenders absolute undecidability results, that there are meaningful questions of mathematics answers to which depend essentially upon the outlook adopted. The Idealist answers the questions one way, the Empiricist another. Since no conclusive mathematical consideration will ever decide between the two, such questions constitute undecidable problems whose solutions will remain forever outside the range of our mathematical abilities. A main point of *General Function Theory* was that one, if not the premiere, such question is the existence of limits for bounded, monotonically increasing sequences. Du Bois-Reymond wrote,

> The solution of the riddle [e.g., that of limits] is, if I am correct, that it is and will always remain a riddle. The simplest expression of the riddle appears to be a psychological one. The most extensive observation of our thought processes and their relation to perception leads us ineluctably to the conclusion that there are two distinct means of conception sharing the same right to count as foundational in the exact sciences, since neither of the two produces results that are disconfirmable, at least when we restrict ourselves to pure mathematics. … These two methods of representation I name, in keeping with the standard nomenclature, … *Idealism* and *Empiricism*. ([11]: 2–3)

For Paul du Bois-Reymond, who advanced his own account of the infinitely small, the question "Does the continuum contain infinitesimals?" was equally pressing. His Idealist argued, along vaguely realist lines, that infinitesimal real quantities do exist and are required for the completeness of the real numbers. His Empiricist stood in opposition, insisting that we have every reason to believe that real infinitesimals are figments of the imagination, in no way required for a satisfactory higher mathematics. Du Bois-Reymond maintained that we will never find grounds sufficiently solid for preferring one position over the other on this issue. In this, as in the matter of limits, our mathematics is and will always remain incomplete.

3. Du Bois-Reymond's Philosophy and Hilbert's Program

Paul du Bois-Reymond's outlook on mathematics included views one would naturally expect in a position to which the Hilbert Program was to be a reply. He and Hilbert would have agreed that, on a proper metaphysics of proof, proofs are seen to be arrays of representations. Du Bois-Reymond held that mathematical thought is a rule-governed manipulation of representations and not, as the intuitionists maintained, a direct intellectual engagement with mathematical objects. He believed the mind to be populated by representations or *Vorstellungen* that are abstract to various degrees. At a level close to perception, representations are extracted from perceptions or intuitions and stand for perceived or intuited objects much as a Xerox copy can stand for its

original. One might call these 'object-representations.' Du Bois-Reymond allowed that there are also representations derived from no real perception but supplied by the mind itself. He called these *Wort-Vorstellungen*, word-representations. In this case, there is no pictorial image but only a word to tag a concept. Our representations for the limits of sequences of rational numbers would often be of this sort. For du Bois-Reymond, proofs were to be combinations of representations satisfying certain mathematical requirements. In the introduction to *General Function Theory* we find, "So much is certainly clear: a proof has to connect either a representation that is already available at the start or the common content of a class of representations, known as a concept, with a new, to be grasped or proven, final representation via a connected chain of representations" ([11]: 11).

Second, for du Bois-Reymond as for Hilbert, a successful proof had to be finite in length and completely surveyable: "For a proof, as with an explanation, is, at bottom and generally speaking, the production of a logically satisfying sequence of representations linking one representation, which engages our concern, to such representations that do not disturb our peace" ([11]: 111). Du Bois-Reymond's analyses of various arguments for the existence of limits, set out by the Idealist in *General Function Theory* ([11]: 61ff), presuppose the finiteness of proofs. For example, the Idealist criticized the standard proof, proceeding by repeated interval division, for the least upper bound property of the real numbers, namely, that every bounded, strictly increasing sequence of real numbers approaches a limit, its least upper bound. As an attempted proof, as a connected representational chain transforming representations for premises into those for conclusions in a finite number of intermediate steps, the procedure of interval division fails, or so reasoned the Idealist. Du Bois-Reymond had the Idealist complain that, in the end, the division procedure cannot produce, from starting object-representations for an interval and a sequence, an object-representation for a single dimensionless limiting point. There can be only finitely many steps in the chain of representations that comprise the proof, he argued. However, after finitely many steps, at most a finite number of divisions of the original interval, enclosing the tail of the sequence, can be made. At a finite stage in the division process, a stage at which the purported proof has to terminate, one is left with, at best, an object-representation of a rational interval of nonzero length (perhaps a visualized line segment) and not the visualized point or dot representing the unique real number that the proof's conclusion requires. "On the basis of our assumptions, we can certainly keep reducing the length of the interval without limit. This is, however, a process which alters nothing in the nature of our representations. Large or small, the interval ... remains always an interval between two rational points" ([11]: 61). Whether defensible or not, this line of thought would make no sense unless the author of *General Function Theory*, in the guise of the Idealist, required proofs to consist of representations and be finite in length.

Third, it would not be wholly anachronistic to ask if, with his *Infinitärcalcül*, Paul du Bois-Reymond thought himself to have constructed a nonstandard continuum. Of course, the contemporary ideas of nonstandard model and of abstract languages sepa-

rate from their varied interpretations were not then available. Du Bois-Reymond held that the properties of his domain of *infinities* or *infinite orders* (not to be confused with Cantor's arithmetic of infinite cardinals, a development wholly distinct) agreed to an extent with those generally supposed to hold among the standard real numbers. Several features of the standard domain capturable by geometrical object-representations, e.g., density, are manifested in the nonstandard number continuum as well. However, the two continua may not have all analytically discernible features in common. Using diagonalization, du Bois-Reymond proved that the infinite orders definitely do not possess the least upper bound property. The domain of orders contained infinitesimals, as du Bois-Reymond was also able to show.

To a first approximation, the face-off between du Bois-Reymond's two characters, Idealist and Empiricist, was a register of divergent attitudes toward the existence of infinitesimals. The Idealist championed a vision on which a nonstandard number continuum forms the true structure underlying the continuum of geometry, a structure open to the analytical intellect but revealed only partially, in glimpses often fleeting and misleading, to geometrical intuition. The Idealist's nonstandard continuum is an ideal, or series of ideals, posited by the mind and inserted to fill gaps in the geometric continuum where limits ought to lie. By contrast, the Empiricist was the champion of a real number system exhausted by intuition; his mathematical empiricism was to be a system "of complete renunciation" ([11]: 3). The Empiricist renounced infinitesimals, restricting himself to intuited real magnitudes. He denied outright that the Idealist's analytical machinery is at work behind the scenes erected by object-representations as a fitting backdrop to mathematical analysis.

Du Bois-Reymond's division of mathematical thought into empiricism and idealism coincided to a large degree with a division into its finitary and infinitary aspects, the first of which maintaining a priority over the second. The Idealist believed the concept of magnitudes infinitely large and that of magnitudes infinitely small to make sense and to be exemplified in reality; the empiricist denied sense and exemplification to both. Du Bois-Reymond had his Empiricist exclaim, "For the construction of mathematics, the finite suffices" ([11]: 146). The continuum of the idealist was to be uncountable in magnitude but that of the empiricist only potentially infinite. It is essential to remember that, in the writings of du Bois-Reymond, Idealist and Empiricist were not so much representatives of different foundational schools, e.g., logicism or constructivism, as discriminable voices in one and the same mathematical consciousness. Every mathematician, du Bois-Reymond thought, sometimes reasons as the Idealist and sometimes as the Empiricist, much as, in Hilbert's scheme, finitism would reign in metamathematical heaven at the same time that infinitary, classical analysis held sway over mathematical earth.

Further, since du Bois-Reymond believed that the continuum of the Empiricist agreed with that of the Idealist in all suitably elementary geometrical and analytical respects, no object-representation and no mathematical datum would ever distinguish between the opposing visions by confirming one and disconfirming the other. The Idealist will never, unless he commits a mathematical error, claim anything to be a

mathematical fact of an elementary character that the Empiricist will not be able to accept. Despite the threat of anachronism, one might suggest that idealistic mathematics was to be conservative over the empiricistic when it came to mathematics acceptable to the latter, much as Hilbert's infinitary higher arithmetic was to be conservative over its finitistic fragment. In consequence, empiricism was to hold, in its strict adherence to geometrical intuition, a manner of priority over idealism. The knowledge it delivers is assured of greater certainty, since, according to du Bois-Reymond, the appeal to geometrical object-representation is either innate for humans or acquired very early in life, while the unfettered analytical thought and word-representations of the Idealist are not. Paul du Bois-Reymond took the results of idealistic real analysis to be a belated adjustment to some original geometrical data: "Accordingly, we can glimpse, in the intercalation of irrational numbers among the rationals, only a retrospective adaptation of the intellective number concept to the concept of geometrical magnitude, which is either innate or acquired in infancy" ([10]: 150). Because Idealist and Empiricist were to agree on all the relevant data, the Idealist could look to the future mathematical progress of the Empiricist as offering further support for his own infinitary investigations ([11]: 148).

Fourth, in the posthumous monograph [12], du Bois-Reymond argued that infinitary mathematical claims are generally ideal in that their terms need not denote anything given by an object-representation drawn from a perceived or intuited realm. Instead, the items represented by those terms should be viewed as symbolic limits added to the finitistic or empiricistic sector in order to round it out. Apart from those magnitudes like $\sqrt{2}$ associated with the visible result of a geometrical construction, irrational numbers are not allied with object-representations but only with word-representations that correspond to nothing in intuited mathematical reality: "The sequence of contentual representations of exactness has as its end result [its limit] a word for something unrepresentable" ([12]: 80). A few pages later in the same work, we find him asserting, "The concept of the infinite that turns up here is the most significant of the unrepresentable word-representations, the foremost idealistic concept, because it is best attached to the question of our conception of the existence of the ideal" ([12]: 85–86). In *General Function Theory* [11], any magnitude that is purely idealistic is nothing but a symbol:

> All the mathematical magnitudes we have introduced so far can be located in the realm of the perceptual world, either mental or non-mental. We wish to call such magnitudes real. There are also magnitudes created by the human thought process and which stand outside any direct relation to the perceptual world. Logical processes for achieving combinations of symbols, one of the favorite activities of the human spirit, lead to certain symbols which are, in mathematics, also called magnitudes, and which serve to assemble into a single sign a mathematical conclusion that frequently recurs. ([11]: 38)

Incidentally, this quotation helps underscore the extent to which du Bois-Reymond considered empiricism to hold priority over idealism. The Empiricist limited his mathematical considerations to those magnitudes which are, as here explained, real. The Idealist did not.

Fifth, as earlier noted, Paul du Bois-Reymond explicitly refused any principle of solvability for mathematical problems. This refusal followed as an immediate corollary to his characterization of the Idealist/Empiricist dualism. The former believed in a continuum that contains, in addition to ordinary magnitudes, infinite and infinitesimal numbers. The latter staunchly maintained that none of these purely analytical adjustments really exists. According to du Bois-Reymond, no completely convincing mathematical demonstration, no mathematical datum, no scientifically respectable argument will be sufficient to prove either the truth or the falsity of the Idealist's claims to the satisfaction of all. The truth of the statement "infinitesimal magnitudes actually exist," crucial to the Idealist's worldview, cannot be decided. It will never be proved; it will never be disproved. There can be no final scientific determination of the structure of the continuum. An examination of the text of Hilbert's Problems Address suggests that his Axiom of Solvability, there enunciated, might well have been intended to exorcize this particular spectre of incompleteness [21].

Finally, Hilbert and Brouwer endorsed, in the strongest terms, the autonomy of mathematics. Du Bois-Reymond rejected it in terms no less vigorous; he insisted that mathematics is not an independent science, but can live only in symbiosis, its foundation shored up by a nonmathematical discipline. He wrote, "As was to be suspected and as we soon recognize, the [foundational] difficulty of the concept of limit is not of a mathematical nature. If it were, it would have been dealt with long ago. The difficulty is really rooted in the simplest constituents of our thinking, the representations" ([11]: 2). The discipline underlying mathematics was to be the proper study of representations. For du Bois-Reymond, that study was the physiological psychology that so prospered in Germany during the 19th Century, among whose foremost representatives had been Johannes Müller, Hermann von Helmholtz and his older brother Emil. In the first section of *General Function Theory*, that devoted to the analysis of magnitude, the authorities most often cited seem to be such physiologists as Müller and Gustav Fechner. Paul du Bois-Reymond believed that physiology will reveal to us by experiment the nature and extent of represented magnitude and, on that experimental basis, the foundations of mathematics could be erected. Only in this way, by a scientific determination of what humans can actually perceive, can the crucial foundational distinction between the Empiricist, who restricts himself to magnitudes fully captured by geometrical object-representation, and the Idealist, who is willing to countenance unperceivable and unintuitable magnitudes, be made with full assurance. The largest feature on du Bois-Reymond's map of mathematical thought, the final treaty-line between empiricism and idealism, was to be drawn by the hand of physiology, not by that of mathematics.

4. Conclusion

Both Hilbert's Program and the Project in which it played a part took aim at a number of dialectical targets; they were not zeroed-in exclusively on the large target offered by the intuitionists. Equally in his sights was Emil du Bois-Reymond's *Ignorabimus*, a grand agnostic scheme and, so, a foil well suited to the broad optimism and wide epistemological sweep intended for the Project. And there were the skeptical philosophy and nonstandard mathematics of Emil's brother, Paul, a combination better placed philosophically to afford a natural opponent to Hilbertism than any intuitionism from Amsterdam. To put the naturalness of that opposition on clear display was the main goal of this writing. No one would doubt that we can gain true and lasting insight into the views of Hilbert and his allies only by knowing, and knowing thoroughly, the ideas of Hilbert's intellectual enemies. Hence, without a due appreciation of the ideas of the du Bois-Reymond brothers and an accurate measure of the powerful influence those ideas exerted on thinkers of their day, there can be no full, rich understanding of the Hilbert Program, not merely in its first beginnings but also in its argumentative ends, forthcoming in our day.

Acknowledgement. I am grateful to Christian Thiel and Peter Bernhard of the University of Erlangen, as well as to Volker Peckhaus of the University of Paderborn, for allowing me to present preliminary versions of this essay to the Colloquium Logico-Philosophicum at Erlangen. I wish to thank Bernd Buldt, Volker Halbach and their colleagues in the Logik Forschergruppe at the University of Konstanz, especially Ulf Friedrichsdorf and Max Urchs, for comments and suggestions on a forerunner of this paper presented there. Bernd Buldt was also more than generous in providing detailed comments on subsequent versions. Essential help was given me by Gerhard Betsch, Mic Detlefsen, James Hardy, Reinhard Kahle, Stuart Mackenzie, Jeremy McCrary and Marianne Moberg-Blauert. Last but not least, I thank Steve Forrest, Robert Noel and Paula Patton of the Indiana University Library in Swain Hall for their patience and determination in tracking down works by 19th Century mathematicians and philosophers now sadly forgotten.

References

[1] Brouwer, Luitzen E. J.: 1927. Über Definitionsbereiche von Funktionen. [On domains of definition for functions.][1] *Mathematische Annalen* 97: 60–75.

[2] Brouwer, Luitzen E. J.: 1928. Intuitionistische Betrachtungen über den Formalismus. [Intuitionistic reflections on formalism.] *Koninklijke Akademie van wetenschappen te Amsterdam. Proceedings of the Section of Sciences*. [The Royal Academy of Sciences in Amsterdam.] 31: 374–379.

[1] Unless otherwise noted, translations from the German are my own.

[3] Brouwer, Luitzen E. J.: 1952. Historical background, principles and methods of intuitionism. In: W. Ewald (tr. and ed.), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics. vol. II*, Oxford: Clarendon Press, 1197–1207.

[4] Brouwer, Luitzen E. J.: 1992. Intuitionismus. [Intuitionism.] Edited by D. van Dalen, Mannheim: Bibliographisches Institut und F. A. Brockhaus AG., 161pp. (Contains an edited text of Brouwer's 1927 Berlin lectures together with Brouwer's *Theory of Real Functions*, both previously unpublished.)

[5] Browder, Felix E. (ed.): 1976. *Mathematical Developments arising from the Hilbert Problems*. Proceedings of Symposia in Pure Mathematics. vol. XXVIII. Providence, RI: American Mathematical Society, xii+628. I have amended the translation of Hilbert's 1900 lecture in the light of [18].

[6] Cantor, Georg: 1891. Über eine elementare Frage der Mannigfaltigkeitslehre. [On an elementary question of set theory.] *Jahresbericht der Deutschen Mathematiker-Vereinigung*. [Annual Report of the German Mathematical Association.] Erster Band: 75–78.

[7] Detlefsen, Michael: 1986. *Hilbert's Program. An Essay on Mathematical Instrumentalism*. Dordrecht: Reidel, xiv+186.

[8] Du Bois-Reymond, Emil: 1886. Über die Grenzen des Naturerkennens. Reden von Emil du Bois-Reymond. Erste Folge. [On the limits of our knowledge of nature. Addresses of Emil du Bois-Reymond.] Leipzig: Verlag von Veit and Comp., viii+550.

[9] Du Bois-Reymond, Paul: 1875. Über asymptotische Werte, infinitäre Approximationen und infinitäre Auflösungen von Gleichungen. [On asymptotic values, infinitary approximations and infinitary solutions of equations.] *Mathematische Annalen* 8: 363–414.

[10] Du Bois-Reymond, Paul: 1877. Über die Paradoxa des Infinitärcalcüls. [On the paradoxes of the infinitary calculus.] *Mathematische Annalen* 10: 149–167.

[11] Du Bois-Reymond, Paul: 1882. *Die allgemeine Functionentheorie*. [General Function Theory.] Tübingen: H. Laupp, xiv+292.

[12] Du Bois-Reymond, Paul: 1966. *Über die Grundlagen der Erkenntnis in den exakten Wissenschaften*. [On the Foundations of Knowledge in the Exact Sciences.] Sonderausgabe. Darmstadt: Wissenschaftliche Buchgesellschaft, vi + 130.

[13] Felscher, Walter: 1978. *Naive Mengen und abstrakte Zahlen II. Algebraische und reelle Zahlen*. [Naive Sets and Abstract Numbers II. Algebraic and Real Numbers.] Mannheim: Bibliographisches Institut Wissenschaftsverlag, 195.

[14] Geier, Manfred: 1992. *Der Wiener Kreis*. [The Vienna Circle.] Reinbek bei Hamburg: Rowohlt, 157.

[15] Hardy, Godfrey H.: 1954. *Orders of Infinity: The 'Infinitärcalcül' of Paul du Bois-Reymond*. Cambridge Tracts in Mathematics and Mathematical Physics 12. Cambridge: Cambridge University Press, 77.

[16] Hilbert, David: 1919–1920. *Natur und mathematisches Erkennen*. [Nature and Mathematical Knowledge.] Berlin: BirkhäuserVerlag, 1992, xiv + 101.

[17] Hilbert, David: 1922. The new grounding of mathematics. In: W. Ewald (tr. and ed.), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics. vol. II*. Oxford: Clarendon Press, 1996, 1115–1134.

[18] Hilbert, David: 1935. *David Hilbert. Gesammelte Abhandlungen*. Dritter Band. [Collected Works. vol. 3] Berlin: Springer, vii+435.

[19] Hilbert, David: 1964. On the infinite. In: P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 134–151. (A translation of Hilbert's 1925 address to the Westphalian Mathematical Society as printed in *Mathematische Annalen*.)

[20] Hobson, Ernest W.: 1907. *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*. Cambridge, UK: Cambridge University Press, xv+772.

[21] McCarty, David C.: 2004. Problems and Riddles: Hilbert and the du Bois-Reymonds. *Synthese. Special Issue for GAP 2000*. To appear.

[22] Pringsheim, Alfred: 1898–1904. Irrationalzahlen und Konvergenz unendlicher Prozesse. [Irrational numbers and the convergence of infinite processes.] In: W. F. Meyer (ed.), *Encyklopädie der mathematischen Wissenschaften. Erster Band in zwei Teilen. Arithmetik und Algebra*. [Encyclopedia of the Mathematical Sciences. First Volume in Two Parts. Arithmetic and Analysis.] Leipzig: Teubner, 47–146.

[23] Reid, Constance: 1970. *Hilbert*. Berlin: Springer, xi+290.

[24] Sieg, Wilfried: 1999. Hilbert's Programs: 1917–1922. *The Bulletin of Symbolic Logic* 5: 1–44.

[25] Stein, Edith: 1987. *Essays on Woman. The Collected Works of Edith Stein. vol. 2*. Translated by Freda Mary Oben. Washington: ICS Publications, ix+290.

[26] Troelstra, Anne S. and Dirk van Dalen: 1988. *Constructivism in Mathematics: An Introduction. vol. II*. Amsterdam: North-Holland, xvii+879+LII.

[27] Webb, Judson C.: 1980. *Mechanism, Mentalism, and Metamathematics. An Essay on Finitism*. Dordrecht: Reidel, xiii+277.

Indiana University
Bloomington, IN 47405-1006
USA

E-mail: dmccarty@indiana.edu

Russell's Paradox and Hilbert's (much Forgotten) View of Set Theory

Jan Mycielski

Abstract. We try to explain the convictions of mathematicians that the axioms of Russell's theory of types and of ZFC are consistent. We claim that this is an inductive inference (a generalization) based on mental experience, the experience of a certain physical constructive process occurring in the brains of those who read with understanding the axioms of ZFC. Thus Con(ZFC) is a fairly well established scientific claim or prediction about a physical process. We describe this process by means of concepts introduced by Hilbert, Poincaré and Skolem, and we add a number of remarks which we believe to be in the spirit of Russell's rationalism.

1. We all know the message of Russell's famous letter to Frege: The class of all sets which are not members of themselves is not a set.

But why should it be a set? And, since it is not, can we develop a clear view of sets (in our imaginations) which shows in a palpable way that this class is not a set?

These questions have been answered almost one hundred years ago, and the answer is: sets are like imagined boxes intended to contain other boxes. Of course we cannot enclose a box into itself. And hence the class of all boxes cannot be put into one box. Thus the class of Russell is a superbox, a container designed to contain all boxes but no superbox.

Henri Poincaré said it differently. He said, for example, that a line is not a set of points, although we can construct many points on it. And the Aristotelian concept of potentially infinite sets also suggests the view that sets are some kind of imagined boxes waiting to be filled up.

David Hilbert in 1904 [3] used still different words. He wrote that sets are thought-objects which can be imagined prior to their elements. [At the request of the referee who asked what is a thought-object let me add: I understand it to be a thought about an object which may exist or not. Thus it is an electrochemical event in the brain or/and its record in the memory. In particular it is a physical thing in space-time. Of course it is difficult to characterise any physical phenomena. But we have the ability to recognize thoughts as identical or different, just as we have the ability to recognize a silent lightning from a thunderous one. Hence I understand Hilbert's words as follows: mathematicians imagine many sets which do not exist, but their thoughts about sets do exist and they can arise prior to the thoughts of most elements in those sets. Moreover, in 1923 [4], he described to some extent the algorithm creating those thoughts, see below section **3**.]

Georg Cantor used the term definite sets, to stress the difference between sets and proper classes. I believe that he must have thought about sets and classes in a similar way, that is as boxes and containers (in spite of his insistence upon actual infinity), because this is the only way I can actualize them in my imagination.

Still, even today, the word set is a bit confusing to some philosophers who are not used to set theory. In fact some are perplexed by Russell's paradox and think that Frege's inconsistent set existence schema is intuitive. It would be better if set theory teachers (and books on set theory) told at the beginning that it is essential to view the universe of sets as a container intended to contain boxes intended for other boxes and one of them intended to remain empty. Of course the set-brackets {...} suggest this view of sets, but this notation should be explained to those students who meet it for the first time.

[However, let me defend Frege. When we construct axiomatic theories it is natural to start with axioms as strong and as simple as possible (even if one does not see through very well). Then, if inconsistencies are discovered, of course one has to prune down the axioms. Thus it was natural that Frege started with the simplest and strongest set existence axiom which occurred to him. Something similar happened when H. Steinhaus proposed the axiom of determinacy. His original form was too strong and I trimmed it down (then we published it together). But I had also another stronger version and D. Scott showed that it is inconsistent; see [6], p. 219. And again when William Reinhardt proposed his axioms of large cardinals K. Kunen showed that some of them were too strong and inconsistent. And recently, some axioms which imply GCH occurred to me and H. Woodin showed that some of them were inconsistent with each other (the weaker versions which still appear to be consistent also imply GCH and seem interesting; see [9]). In this kind of work one has to take some risks.]

The view of sets and classes as boxes and containers points out to the constructive process in human imaginations creating set theory. It suggests the thesis that the Skolem Paradox of existence of countable models of ZFC is not at all a paradox but a preliminary mathematical description of this process. Hilbert's ontology of mathematics has been understood as follows: Pure mathematics is a tale about nothing (i.e., about something completely imaginary). This is called *formalism*. It is unsatisfying, since it does not explain why mathematics (the set theory ZFC) appears to be consistent, and why it appears to be a discovery and not a pure invention. I think that his view (at least in 1904) was deeper, namely that pure mathematics is a description (however indirect or metaphorical) of a growing finite structure of thought-objects actually constructed in human imaginations, and that set theory is a straightforward extension of human natural logic, of the ability to classify things into sets, and of the logic and the potential for creating language given to us by natural evolution. Perhaps an essential part of this view goes back to Aristotle (I am not familiar enough with ancient mathematics to make any firm attributions). This view will be called here *rationalism*. Notice that it is free of Platonist ontological assumptions. It explains the belief in the consistency of ZFC as an inductive inference based on a mental experiment. It explains also such works of fiction other than mathematics where something

sufficiently concrete has been imagined such that it can be described in many ways and in any common language. Later (in 1923) Hilbert proposed a formalism (his ε-symbols) for a direct (non-metaphorical) description of such growing finite structures. Those structures were also described in 1920 and 1922 in mathematical terms (informally) by Thoralf Skolem; see Sections **3** and **4** for more details.

Thus rationalism offers a biological explanation (and description) of the nature of pure mathematics and of all sufficiently concrete imaginary structures, taking them to be physical processes and their records in human brains. In the remainder of this lecture I will try to explain rationalism more fully and I will add some remarks (or impressions) about its history. Of course this account is only a sketch which is far from completeness.

2. Let me begin with some events from my early education.

I asked my high-school teacher if mathematics is consistent. He said that he is not sure since there are things called paradoxes in set theory. But he said that some have worried about that question and this stimulated the writing of some books. He lent me one, a semiformal exposition of what was then called theoretical arithmetic. It was full of close-knit proofs of obvious facts, the whole thing pedantic and boring, and absolutely inconclusive relative to my question.

Of course I learned later that Hilbert asked the same question, and that Gödel has shown that in a sense we do not and cannot know the answer.

But, in spite of this, I had an experience which I can repeat at will: When reading the formulas expressing the axioms of set theory, I form in my imagination an approximate picture of the universe of sets which is so concrete that it convinces me that this system of axioms is consistent.

How to analyse this experience? What is happening in my imagination? I believe that I construct mentally a little finite segment of the Skolem hull of sets generated by the empty set by means of the operations suggested by the axioms. This construction appears so regular, in a sense periodic, that I believe that it can be continued forever. The injection of the infinite set is beautiful, no obstacle whatsoever. (See also below, Section **3**.) The physical prediction that ZFC will never be found inconsistent is equivalent to the prediction that these finite segments of Skolem hulls can be extended for as long as we can and wish to do it (even with the help of computers). [A more detailed mathematical definition of these finite structures arising in human imaginations and supporting the feeling of consistency and concreteness of mathematical theories is given in [8].]

I do not want to be disrespectful to those who are believing Platonists, but it seems to me that this experience in mental construction has nothing to do with seeing a Platonist universe of sets. It derives from a certain regularity of ZFC. If I had been exposed e.g., to Quine's NF I believe I would not have had such an experience since the Quinean membership relation has a loop $V \in V$ and so the finite segments of Skolem hulls for NF are not that simple. I was relating a personal experience; perhaps all of you had this experience (if not with ZFC, then with PA)?

[A more general observation may be of interest: The fact that we believe the consistency of ZFC is caused by a mental mechanism M which turns into convictions this (and many many other) inductive generalizations, based on mental or external experiences. M acts in such a way that it turns into convictions any description or theory which appears to be true (that is to agree with facts in some intended measure) and is the simplest (or a disjunction of the simplest) among those descriptions which we know and which seem to have this agreement. [To accept the simplest among the fullest descriptions is often called *inference from the best explanation*.] Of course there are two ways of checking such an agreement: we learned that description from a source which we can trust, or we made ourselves an observation or an experiment which confirms it strongly enough. Moreover, M acts independently of our will. In particular it is difficult or impossible to loose our belief unless we forget it or we learn a better description (i.e., one which is simpler and/or appears to satisfy that agreement in a larger degree). Notice that M is an attribute of good intelligence, and that it counters any general form of scepticism. Let me add that as soon as evolution gave us the power to form freely complex mental models of reality and the language to communicate them, it had to give us also an instinctive preference for the simplest such models, i.e., those which are the easiest to remember and to communicate from scratch. And M is a perfected form of this instinct. I believe that there exists also another reason for that preference: the simplest theories which appear to be true have a better chance to be true (in their intended measures). In the practice of daily life and in science this preference is always obeyed.]

Thus the questions, why mathematics appears to be consistent, why it is so concrete or reconstructible in our minds, are all answered. It is all grounded in the simplicity of the axioms of ZFC and the human ability to carry a mental construction of appropriate finite structures when exposed to those axioms. (Although it is possible that some like Zermelo himself may first imagine the structures and then express the axioms.)

Still, I do not think that ZFC has any absolute position. Most extensions of ZFC with axioms of existence of large cardinals even some very strong ones (like Vopenka's axiom or Reinhard's n-extendible cardinal existence axioms), are easily incorporated in the above mental construction which yields in the same way the conviction that they are consistent. But, of course, as the axioms become stronger the conviction becomes weaker. I believe that our minds have some subjective estimates of the probabilities of consistency of those theories; however those values are not easy to measure, they are present only in our subconscious knowledge and they vary with time and experience. It is somewhat like evaluation of positions in chess. (Concerning our evaluating knowledge see below Section **5(e)**.) More generally, all our beliefs may have such probabilities in our minds. (Thus the concept of a conviction used above is a simplification. Rather than creating convictions M attaches probabilities of consistency (and of truth when this concept applies) to theories. But we are able to observe clearly the presence of these probabilities in our minds only when they are close to 1 or close to 0.)

Of course one should remember that the mental constructions supporting the consistency of ZFC and its natural extensions are not full proofs. For example it seems to me that if somebody is not familiar with set theory, and he is exposed to ZFC + AD (where AD is the Axiom of Determinacy), he may have the same positive experience. AD may appear to him as a natural generalization of a fact on finite games of perfect information. Indeed the clash between AD and the Axiom of Choice is not immediately visible since it requires a longer proof. At least I can tell that if I omit the axiom of choice, that is I consider the theory ZF + AD, then that mental experience supporting consistency happens in my mind. And I believe that whenever we consider a new theory, at the beginning we have only that mental constructive method and no Platonic telescope to see anything.

3. In 1951 the logician J. Słupecki told me: I understand free variables. They are places for the substitution of more concrete terms. But I do not understand quantifiers since often they refer directly to some actually infinite universes which I do not believe to exist. (Like a number of Polish logicians of that time Slupecki called himself a nominalist; so did Kotarbiński and Tarski. They were influenced by K. Twardowski. See also [8].)

But there was an excellent answer to his question which Hilbert gave in 1923 [4] which I did not know in 1951 (it is possible that Slupecki knew it and wanted only to stimulate his student): Hilbert's ε-symbols eliminate quantifiers from the basic formalism of logic. Quantifiers become abbreviations or defined symbols. Namely,

$$\exists y\varphi(x, y) \leftrightarrow \varphi(x, \varepsilon y\varphi(x)), \quad \text{and} \quad \forall y\varphi(x, y) \leftrightarrow \varphi(x, \varepsilon y(\neg\varphi)(x)).$$

In fact, if we add to first-order logic without quantifiers the ε-operator, and Hilbert's axiom schema (H):

$$\varphi(x, y) \rightarrow \varphi(x, \varepsilon y\varphi(x)),$$

and the above definitions, then one can prove the usual axiom schemata on quantifiers.

Even variables can be eliminated, since the above formulas can be interpreted as *axiom schemata* of logic, where the free occurrences of the variables x and y are understood to be arbitrary *constants* (i.e., terms without variables). (By *axiom schemata* we mean certain rules which yield theorems.)

Thus both quantifiers and free variables in any statements are explained via Hilbert's ε-symbols as schemata for formulas without any variables free or bound. In this way, formulas with free variables can be understood as rules for forming quantifier-free sentences (i.e., formulas without any variables) and those sentences do not refer to any objects except those whose names appear in them, in particular they do not refer to universes. Of course if y is the only free variable in φ, then the term $\varepsilon y\varphi$ is a constant. This constant can be called a *general object*, since its only intended property is expressed by the schema (H), i.e., it is maximizing the truth value of φ. (It is useful to generalize all this in such a way that we understand x and y to be any finite sequences of variables, and $\varepsilon y\varphi(x)$ as a sequence of $|y|$ function symbols each having

$|x|$ variable places, where $|s|$ denotes the length of the sequence s.) [Concerning the mental role of variables see also below Section **5(a)**.]

Now notice the following ambivalence of sentences of first-order logic. The *Platonist* understands them as assertions about an ideal (most often actually infinite) structure. The *formalist* takes them to be meaningless strings of symbols. The *rationalist* (me) understands them as assertions about finite structures of constants which are named in those sentences. Also, depending on the context, the rationalist understands formulas with free variables as names of relations or as rules for producing sentences. And he understands terms with variables as rules for producing constants (without claiming that he knows or can always decide the truth values of equality or other relations of his language when they are applied to those constants). Thus all three use the same mathematical language, but the rationalist is interested in the physical structures of thoughts underlying them (in human brains), and for him the view of Platonists appears false and the view of formalists trivial. It is clear that rationalism generalizes such that it explains not only pure mathematics but all works of fiction expressed in common languages, and it explains it all in biological terms such as thought-objects of various kinds, and the calculus of terms, relations and Boolean connectives of the natural human logic.

4. I do not know if Hilbert ever juxtaposed his view of 1904 (that sets are mentally actually constructed thought-objects whose members need not be actually constructed), with his ε-symbols of 1923 (which provide notations for all thought-objects). He does not mention either that the ε-operator can be viewed as a tool for naming Skolem functions (1920–1923, see [16]) nor, as we have explained above, that they provide a language for describing directly that constructive process in our brains which is stimulated by reading the axioms of ZFC (see [8] for a fuller mathematical definition of those imagined structures).

In his later writings Hilbert introduced a confusing distinction between concrete and abstract concepts which does not match his 1904 idea (and is not related to the ε-symbols). Of course we have a distinction between those concepts which are used in applications of mathematics to represent physical objects and processes, and those objects of pure mathematics which have no such uses, see [10]. But this does not seem to be the distinction which he had in mind or at least this is not how his commentators understood it (for example it is understood that individual real numbers are not among his concrete objects, however they have of course direct physical interpretations). It seems that he intended his concrete objects to be those which can be denoted by finite strings of symbols (e.g., integers). But then, according to the rational point of view, he made an ontological error since either this implies that there are mathematical objects which cannot be actually denoted or imagined (Platonism) or that all objects are concrete (and the distinction is useless).

[Perhaps we can rescue Hilbert's distinction saying that a structure consists of concrete objects if we have a system of names for all its individual objects and a decision procedure for equality and the other basic relations when applied to those

names (such a structure is called a *computable structure*). For example the individual objects of number theory (integers), of logic (terms and formulas), and of finite combinatorics (hereditarily finite sets) are concrete, while those of second-order arithmetic (real numbers) are not. But, since I do not know any ontological significance of this concept, I doubt if this was the intention of his distinction.]

5. Finally let me add miscellaneous thoughts about the philosophy of mathematics and even philosophy in general (as mentioned above, the ontology of mathematics presented here has a more general significance). Those remarks will complement some points made earlier, but they are somewhat incomplete. I am emboldened to make them because the previous text seems already diametrically opposite to the vast majority of ontological opinions expressed by philosophers, also to the current opinions of phenomenologists and of postmodernists. Indeed I believe that rationalism (as defined above) puts in doubt the cognitive value of a large part of the philosophical literature. It suggests also that as a rule the literature about the history of philosophy is defective since it is uncritical and hence confusing. Many writers stress dead ends and present them as viable alternatives, rather than trying to disentangle solid contributions to knowledge or to distinguish that which is known from open problems and to illuminate the political or/and religious pressures which may have caused errors. (Perhaps there exists a dubious belief that error is a necessary precursor of knowledge. In fact error often obscures the truth and makes it more difficult to construct knowledge. Ignorance, when explicitly acknowledged, is much more useful.) Then some philosophers try to justify it all claiming that philosophy is incapable of producing any definitive answers to its core problems. This in turn appears to me to be an exaggeration. Moreover, I believe that philosophy is not like pure mathematics. We can know mathematics, as we can know other man-made structures, so metamathematics is a knowledge since it is a mathematical theory of a reality, but pure mathematics is not a knowledge since it talks about imaginary structures. On the other hand the aim of philosophy is to be a part of knowledge. In fact solid philosophical knowledge exists and I hope that rationalism belongs to knowledge. Also, philosophical criticism in science is dearly needed (especially in physics). But philosophers left it to the scientists who are not very good at it. Too often their appreciation of basic unsolved problems is insufficient and they overemphasise algorithmic and technical know-how.

(a) Mathematicians often leave interpretations and applications of their concepts to others, and Hilbert may have omitted anything that appeared too obvious to him. In philosophical matters he would rather write about things with which he had difficulties, than those which were clear to him. Then his commentators could have missed some meanings which he had in mind. Perhaps the idea that all thoughts and memories (and hence all mathematical objects) are physical things in human brains would have been obvious to many scientists at the turn of the nineteenth century.

I believe that this view implies that there are no objective degrees of concreteness of thought-objects (I will return to this point in a moment). However thoughts can be fragile in the sense that they are fleeting and not reproducible, or solid in the

sense that they are remembered and can be explained to others in many ways. More important is the objective distinction between those imagined thought-objects which name and are intended to model other physical objects or processes (such as many geometric figures), and those thought-objects for which there is little or no chance of finding physical significance (such as a well ordering of the real line). This is also discussed in [10]. Thus in my vocabulary concrete means physical, but it does not mean meaningful or representing something real. Thus in pure mathematics we write symbols like 2 or $\prec$ (a well ordering of the real line), and both are symbols for concrete thoughts, in spite of the fact that 2 is definable and it plays an important role in numerous applications (e.g., as a property of objects which we view as consisting of two sub-objects) while $\prec$ is not definable in the language of ZFC (prior to the ε-extension) and it has no intended direct physical interpretations. That is, when a mathematician uses $\prec$ in a proof, $\prec$ is as concrete in his imagination as any other thought-object.

Returning to our discussion of variables, let me amplify once again that they are not only places for substitution (as told by Slupecki). When we use them in calculations and in proofs, they get in our minds the status of thought-objects, objects which can be called *general* and denoted by constant ε-terms. For example consider two expressions

$$x = x,$$
$$\varepsilon y \neg(y = y) = \varepsilon y \neg(y = y).$$

The first can be explained as an abbreviation of the second (which is a more explicit). We do not care about (we abstract) all properties of the object $\varepsilon y \neg(y = y)$ (except that it maximises $\neg(y = y)$, i.e., we can use (H)), so we call it a *general object*. [But free variables are also used in different ways such as in the expression $\neg(x = y)$. Indeed, this expression cannot be understood as a statement but either as a notation for (the thought of) the relation of non-equality, or as a schema for some statements without variables.]

(b) H. Poincaré made the following error: With his rational interpretation of infinite sets and classes as potentially infinite sets and classes (we said boxes and containers) he jumped too hastily to the prediction that Cantor's theory of infinite cardinal numbers will never be an interesting or genuine part of mathematics. Thus he thought that one kind of (presumably too abstract) objects can be a priori banned from mathematics. It appears that he believed also that every good mathematical theory must contribute to natural sciences. But such a requirement is inconsistent with his own statement that *in mathematics to exist is to be imagined and to be free from any known contradiction*. Indeed, according to most of us set theory is a good and even admirable theory.

(c) Let us reflect upon the question what information or what concrete structure of mathematical objects is so valued by the constructivists or the intuitionists and according to them is lost in classical mathematics? See e.g., [1]. According to our rational interpretation if an existential statement is postulated or proved, then the corresponding object must be a concrete thought. (Concrete does not mean fully defined in the usual first-order language, for example we can think of an integer

$\varepsilon y(y = y)$, and we do not wish to decide its parity, thus it remains a general object; or we can think of a pencil, but we do not think about its color or length). This suggests that nothing valuable is lost in classical mathematics.

The only proposal of intuitionists whose motivation is clear to me (although they do not express it in this way) is that mathematics should deal only with computable structures. But I think that this restriction would badly impoverish mathematics. The motivation of their second restriction, namely rejecting proofs using the law of excluded middle, is much less clear to me and it would damage mathematics even more. It is true that this restriction automatically forces the mathematician to prove more, however if the effect of that additional work is not expressed in the theorems it is lost. I believe that it is better to prove stronger theorems in a classical way, than to hide their meaning in their intuitionistic proofs.

[Intuitionistic logic does not agree with the way we organize theories and descriptions of reality. We observe that in the world at large theories and arguments are built such that *p or not-p* is accepted as true without necessarily knowing that *p* is true or knowing that *not-p* is true, and classical mathematical thinking is not any different from common thinking. (Of course in our imagination *p* is neither true nor false, it is a sentential variable, a general object of a certain kind.) Thus human knowledge is organized on the basis of classical logic and not intuitionistic logic; see also [5], [8], [10].]

(d) All the discussions of truth and existence in mathematics in the collection [2] propose or assume theories which are unconvincing since those theories are more complicated than rationalism. Thus a few words on the concepts of truth and existence are in order. As we know, *truth* (in its primary meaning) is an agreement between sentences and the reality, and an adequate mathematical theory of this relation was given by Tarski. [This theory inspired model theory, but the latter went far beyond the original epistemological motivation of Tarski.] According to this primary sense, since pure mathematics talks about imaginary objects, there is no truth in it, and the word *true* (when we talk about mathematical statements or theories) has a different meaning than in normal parlance. It seems so easy to explain this meaning that I do not understand why philosophers wrote papers devoted to this issue. Namely, when talking about the truth of mathematical statements we are in the realm of metamathematics, see also [7]. Thus *A is true* means *A has been assumed or proved from axioms and definitions which are clear from the context*. If we challenge the statement *A is true* we question the existence of such a proof of *A*. Spoken and written mathematics does not describe directly the structure of mathematical thoughts. We can say that it is a true but metaphorical description of our finite thought-structures. A direct description, i.e., a translation of mathematics into the variable-free language defined in Section **3**, would be too long for human uses. [But I think that it will become useful for computers since for them such translations of the usual expressions are not be too long while the rules of proof of the variable-free language are more uniform.]

Also in pure model theory truth or satisfaction do not have any physical meanings. Such meanings appear only when model theory is applied to explain the property *true*

of real thoughts or texts describing reality. So once again, *truth can appear only in applications*.

In a similar way the term *exists* means something else in everyday language, in philosophy, and in science, than in mathematics. First consider the trivial sentence: *Everything exists*. Of course the concept of existence used here does not distinguish anything from anything and hence it is useless. So the primary (and useful) meaning of the term exists must be different. In fact its role is to distinguish those thought-objects which are intended to name or correspond to something from those which are purely imaginary (that is without such an intention). But in pure mathematics the meaning of existence is different. It distinguishes thought-objects which appear to be consistent from the inconsistent ones or, in other contexts, *A exists* means that we were able to prove (in a sense clear from the context) the theorem that an A with such and such properties exists. (Inconsistent objects do appear in mathematics on a temporary basis in proofs by *reductio ad absurdum* and in some preliminary attempts to construct consistent objects.)

Mutatis mutandis all of the above applies to all works of fiction. (Concerning physical interpretations of mathematical objects, see also [10].)

(e) It is more difficult to explain or to build an interesting theory of *intuition*, see [11] and references therein. There are striking examples of the importance of intuition: A good chess player uses not only computation but also his intuitive knowledge in order to choose a good move, a mathematician uses intuition to form interesting axioms, definitions, conjectures, and to build proofs. In each case there are too many possible choices and their consequences are too remote to be explored, and one applies one's *evaluating knowledge* (EK), see [8]. The EK which is useful in chess and mathematics is fully learned since there was no evolutionary pressure to give us such facilities. In [8] I divided knowledge into EK and *descriptive knowledge* (DK), and proposed some mathematical definitions and theories of the learning process of EK and DK. Here I will only add that intuition, the mechanism M (see Section 2) and EK are closely related to each other. Indeed the evaluation of the degree of truth of a statement, of the strength of an analogy, or of the simplicity of a theory or a description, are all made on the basis of EK, and they are essential for the action of M. Vice versa M plays an important role in the construction of EK.

Assuming that our mathematical theories of EK and DK (see [8] and references therein) are adequate, the main unsolved problem is to explain how EK and DK influence each other's development and how they collaborate in our minds.

(f) The philosophical significance of proof theory is unclear, since there seems to be no ontological or psychological significance of the proofs of consistency of arithmetic. To me, and perhaps to most of us, like all pure mathematics, those proofs are motivated only by the formal value or beauty of their logical and conceptual structures (unlike, say, Mathematical Physics which has also another goal: to describe physical reality).

(g) Likewise any special ontological significance of primitive recursive arithmetic (PRA) is unclear except for those areas which touch upon computer science or brain science. For example, the idea that PRA constitutes the appropriate theory of concrete

objects fails, if we agree (see **(a)**) that all thought-objects of pure mathematics are equally concrete in human imaginations.

(h) We do not share Hermann Weyl's concern about the lack of scientific content of mathematics, see e.g., [12]. Weyl did not regard as sufficient the artistic or aesthetical motivation of mathematics. As a pure mathematician he did not feel secure and wanted to do applicable mathematics. Unlike him Hilbert felt secure and I think that most of us share Hilbert's feeling. And we are sure of the importance of mathematics since we know that it is a mental art with an extraordinary esthetical quality, that throughout human history art accompanied and often preceded knowledge, and that mathematics is a prerequisite to science. For these reasons the concern of Weyl seems unnatural.

(i) Historians and philosophers divide the philosophy of mathematics into three schools: logicism (Frege, Russell and others), formalism (Hilbert and his school) and intuitionism (Brouwer and his school). As we wrote in [5], and in view of the ontology described in the present paper, this classification seems artificial and confusing. Also philosophers use a number of terms which are derived from the theological disputes in the Middle Ages, and are in conflict with modern natural language (but pleasing to some of them). Let me discuss those points.

First, what is logic? A natural sense of this word is: the ensemble of those methods of correct argument with which nature endowed human beings. (I am aware that postmodernists do not believe in such a thing. They claim that those methods were made by people and not by nature, that they are cultural and not natural. But I do not believe the postmodernists. The agreement between mathematicians and scientists of all races and cultures and the degree to which knowledge is a cumulative construction disproves the postmodernist ideas; see also [13]). If you agree with the above definition of logic, it is reasonable to regard set theory as a part of logic. Indeed all mathematicians know intuitively enough set theoretical constructions and axioms to build the objects and to prove the theorems in their fields of interest. However, many of them do not know those methods and axioms in a conscious way. They have the know-how but not the know-why (Plato pointed out this distinction in a different context). Thus type theory or even set theory and some special tools for using them, appear to be ready in our nature. [Some must think that it is only in the nature of gifted people, since they write books and offer courses for the ungifted ones; books with titles like "Learning to Reason", which are hundreds of pages long and have almost no mathematical content. I doubt if they can ever be useful; in fact I believe that they are harmful since they can kill the interest of their readers. The compactness and the expressive power of mathematics are essential qualities which are important in our culture, and these qualities are completely missed in those books. Even their authors seem to know that those texts are training programs, but they seem unaware that they are extremely inefficient as sources of information and are dangerously counterproductive as initiations to science or mathematics.] Thus I believe that the primitive sense of the word logic encompasses set theory (say ZFC, or some system with urelements). Nevertheless, since Gödel's discovery of the Completeness Theorem for first-order logic (FOL), in mathematics, the term logic is often restricted to mean FOL. And of

course, after restricting it in this way, one can claim that the program of logicism (of Frege, Russell and Whitehead) of deriving mathematics from logic, is not doable. But, if we return to the old more inclusive sense of the term then this program is doable and it has been fully completed.

And then, formalism (understood as a program) is really the same as logicism, whence the distinction between them rests on the above terminological convention; it vanishes if we adopt the wider sense of the concept of logic. (I am not aware of any significant disagreements between logicists and formalists, unless one takes the Platonism of Frege to be a part of logicism—Russell would have rejected such an inclusion). Thus I think that this distinction is not interesting. Both groups of people tried and (on the basis of the work of Dedekind, Cantor, Zermelo and Skolem) succeeded in constructing a simple and formal (i.e., mathematical) account of mathematics.

Now, the name 'formalism' for the philosophy of Hilbert (apparently invented by Brouwer) should not be used. First, this term carries a pejorative overtone, second, it is misleading. Indeed the mental experience supporting the consistency of ZFC which was discussed in Section **2** is real and, for example, the distinction between ZFC and NF mentioned there is also real. Thus NF could be called formalistic while ZFC intuitively true, and we see a real difference between formalistic and intuitive theories. Moreover, in that mental experience, there exists a significant difference between sentences and objects or structures, and there are many categories of mental objects (the division into the category of truth valued expressions and object valued expressions is only the first step in this analysis, see [8]). And there exists a rich structure in our imaginations and memories which is based on these categories. Hence the structure of mathematical knowledge is far from arbitrary. Hilbert stressed the role of imagination and intuition in mathematics and one cannot attribute to him the view that mathematics is an *arbitrary* game of symbols. Unfortunately as much as I know he did not respond himself to the curious accusation that his philosophy assumes or implies such arbitrariness.

The phenomenon of relatively wide acceptance of biased misnomers like the name 'formalism' for mathematics based on classical logic or for the philosophy of Hilbert, is not isolated. Time and again we see people who twist language in order to win the popular debate. For example the positive overtones of the word realist must delight Platonists who call themselves realists. (They do not seem to be bothered that this contradicts the common meaning of words according to which only an idealist can think that mathematical objects are real, i.e., exist independently of mankind, while a realist knows that they are ideal, i.e., imagined.)

The common meaning of the word rationalism is also rejected by many philosophers and replaced by a contrary meaning, namely they oppose rationalism to empiricism (but in common parlance it is irrational to ignore experiments). I do not think that such uses are motivated only by an outdated medieval tradition. They sound to me (at least sometimes) as furtive attempts to diminish the value of science and reason.

Still on terminology: Poincaré was called a conventionalist. Again I do not know any good justification of this label. It is true that he stressed that the choice of primitive

concepts in a mathematical theory is a matter of convenience, a convention. But this is a well recognized fact and not a philosophical position. He may have erred by not realizing that his version of special relativity is not only a convention but a real improvement of the Galilean theory (since the concept of ether becomes redundant). This was an oversight, but not a philosophical position.

(j) I know only one interesting distinction in the philosophy (ontology) of mathematics namely the distinction between Platonists and rationalists, and this motivates the following remarks.

Some Platonists base their conviction on a feeling that PA is true and that every sentence in the language of PA is true or false. What matters is what do they mean by true? After all rationalists have just the same belief: for every sentence in the language of PA either it or its negation expresses a physical law, i.e., a correct prediction about an appropriate physical process. Platonists say also that they are sure that if set theory is done in other planetary systems it is compatible and perhaps very similar to ours. Again rationalists agree since they think that set theory is founded upon some finite mental experiments and there are universal physical laws which determine the outcomes of those experiments.

Thus the only real difference between rationalists and Platonists are the beliefs of the latter that there is a non-physical Platonic world of mathematical objects, and a human ability to learn certain facts of that world. Rationalism explains mathematics without these assumptions (by accepting the physical nature of mental life, which in the present state of knowledge about the brain is obvious). This Ockhamian economy of rationalism makes it more convincing than Platonism.

As mentioned in Section **2** we have a psychological mechanism M which acts independently of our will and makes us believe the simplest explanations of things which agree with facts. Thus we know why rationalism is convincing and Platonism is not. (Notice that although our minds do not tolerate inconsistency, they are more liberal about truth. Indeed M induces beliefs in some theories which only approximate the reality in some intended degree, a degree which, very often, is not known to the believer. E.g., I believe that each of the three theories of planetary orbits of Copernicus, Kepler and Einstein is true in an intended measure in which it is used today, but I do not know how accurate they are.

(k) In this state of affairs the existence of Platonists in this day and age is puzzling to me as it was puzzling to Tarski (see [8]) and presumably to Russell. But let me try to answer this question. Of course there were very outstanding Platonists, and among them Frege, Zermelo and Gödel. But notice that Frege and Zermelo did their main works before the invention of Skolem functions or Hilbert's ε-symbols, and hence we can assume that they did not see any way to justify the rationalistic theory. It remains to explain the position of Gödel who knew very well those concepts of Skolem and Hilbert (the philosophy of Gödel was taken very seriously by several authors; see e.g., [2], [11], [14]). My explanation is that Gödel never thought seriously about the idea that mathematics is a physical construction going on in human brains. It is possible that he had a mystic non-physical concept of mind, since we know that he believed

that the mind discovers or sees (rather than constructs) the universe of sets. Of course many people, and many mathematicians, think that way. But I propose that this is a consequence of a gap in their philosophical knowledge and a propensity to create a quick fix for that gap, somewhat justified by the fact that this fix has almost no impact on mathematics. [However it does have influence upon the attitude toward the axioms of set theory, and perhaps on the way they are invented, see e.g., [6], [7], [9]. I cannot believe that the illusion that they tell a truth about some real infinite universe could be better than the knowledge that the only reality inspiring and supporting them is a finite constructive process in human brains.]

(l) The defense of rationalism offered by philosophers is weak and their influence on sciences almost nonexistent. In fact, as the biologist E. O. Wilson [15] and the physicists J. Bricmont and A. Sokal [13] explained, the lack of consilience between philosophy and sciences is striking (and sometimes it reaches comic proportions). Let me mention one example of this lack of contact or awareness. The philosophers D. Armstrong, A. Flew, B. Russell, G. Ryle, L. Wittgenstein and many others do not quote nor use in their theories and historical accounts the work of physiologists such as Charles Sherrington or Santiago Ramon y Cajal, or the mathematician A. Turing. This is perplexing since the discovery of the complexity of the brain and the construction of relatively simple universal Turing machines provided a much higher platform for a rational philosophy of the human being (the identity thesis) than the platform of earlier centuries upon which those philosophers appear to be standing.

I think that philosophy of mathematics (and philosophy in general) becomes relevant and true to its original aims only when it constitutes a contribution to knowledge. And, to propose a sufficiently sharp (perhaps even useful) criterion let me repeat that *knowledge consists only of the simplest among the fullest descriptions of reality.*

(m) Finally, concerning the future, I believe that the rational philosophy of mathematics points us toward a better understanding of the structure of mathematical knowledge in human brains, and eventually it will explain how mathematicians invent proofs of conjectures. (I have tried to formulate some preliminary mathematical definitions for such a theory in the lecture [8].)

References

[1] Bishop, Erret and Douglas Bridges: 1985. *Constructive Analysis* Berlin: Springer.

[2] Hart, Wilbur D. (ed.): 1996. *The Philosophy of Mathematics*. Oxford Readings in Philosophy. Oxford: Oxford University Press.

[3] Hilbert, David: 1904. On the foundations of logic and arithmetic. In [16]: 129–138.

[4] Hilbert, David: 1923. Die Logischen Grundlagen der Mathematik. *Math. Annalen* 88: 151–165.

[5] Marek, Viktor W. and Jan Mycielski: 2001. Foundations of mathematics in the twentieth century. *Amer. Math. Monthly* 108: 449–468.

[6] Mycielski, Jan: 1964. On the axiom of determinateness. *Fund. Math.* 53: 205–224.

[7] Mycielski, Jan: 1995. New set-theoretic axioms derived from a lean metamathematics. *J. of Symbolic Logic*: 191–198.

[8] Mycielski, Jan: 2001. On the tension between Tarski's nominalism and his model theory (Definitions for a mathematical model of knowledge). Warsaw: Proceedings of the Tarski Conference, to appear.

[9] Mycielski, Jan: 2003. Axioms which imply GCH. *Fund. Math.* 176: 193–207.

[10] Mycielski, Jan: to appear. Pure mathematics and physical reality (continuity and computability).

[11] Parsons, Charles: 1995. Platonism and mathematical intuition in Kurt Gödel's thought. *The Bulletin of Symbolic Logic* 1: 44–74.

[12] Sieg, Wielfried: 1999. Hilbert's programs: 1917–1922. *The Bulletin of Symbolic Logic* 5: 1–44.

[13] Sokal, Alan and Jean Bricmont: 1998. *Fashionable Nonsense, Postmodern Intellectural's Abuse of Science*. New York: Picador.

[14] Tieszen, Richard: 1998. Gödel's path from the incompleteness theorems (1931) to phenomenology (1961). *The Bulletin of Symbolic Logic* 4: 181–203.

[15] Wilson, Edward O.: 1998. Consilience among the great branches of learning. *Daedalus* 127: 131–149.

[16] van Heijenoort, Jean: 1967. *From Frege to Gödel, A Source Book in Mathematical Logic, 1979–1931*, third printing. Cambridge, MA: Harvard University Press.

Department of Mathematics
Campus Box 395
University of Colorado, Boulder
Boulder, CO 80309-0395
USA

E-mail: jmyciel@euclid.colorado.edu

Objectivity: The Justification for Extrapolation

Shaughan Lavine

Abstract. Set theory can be obtained by extrapolating, in a mathematically precise sense, from the mathematics of indefinitely large finite sets. That extrapolation is, I have argued elsewhere, the origin of and motivation for infinitary set theory. But on what grounds can it be argued that the extrapolation is not merely a technical trick but a justified move from knowledge of the indefinitely large to knowledge of the set-theoretic infinite? The application of mathematics to physics provides those grounds.

The finite mathematics of the indefinitely large, with its context-dependent bounds of availability, provides a natural setting for a theory of measurements. Insofar as a successful theory of the correlations and behavior of some types of physical measurements is construed as a theory of the correlations and behavior of measurements of objective physical quantities, the theory of the objective quantities themselves must be obtained by replacing the context-dependent bounds of measurement by the bounds of "measurement" of the physical quantities imposed by objective external physical reality. Those bounds are fixed once and for all, equal to what exists in the hypothesized objective world. But that change yields the extrapolated, infinitary, version of the measurement theory.

Physics makes use, not only of measurements of values of individual physical quantities at points, but of measurements of values of field quantities throughout a region, that is, of the measurement of functions. Insofar as measurements may take on arbitrary values, the extrapolated mathematical theory will be one that allows arbitrary functions. The need to integrate differential equations then inevitably leads to modern set theory, and the success of extrapolation in that part of mathematics that is applicable to physics provides evidence of its legitimacy as a general method within mathematics.

I have argued [10, 11] that infinitary set theory originated through extrapolation (in a technically precise sense) from a theory in a system without commitments to the actual infinite. That system is what I have called finite mathematics of the indefinitely large, which is centered around mathematical work of Mycielski [16] and Pawlikowski [19].[1]

[1] Mycielski and Pawlikowski associated a theory $\mathrm{Fin}(T)$ with any ordinary theory T. The theory $\mathrm{Fin}(T)$ has a vocabulary obtained by augmenting the vocabulary of T with new unary predicate symbols Ω_p, for all rational subscripts p. The axioms of $\mathrm{Fin}(T)$ include those of T regularly relativized to the Ωs: $(\forall x)(\exists y)\phi$ becomes, for example, $(\forall x)(\Omega_0(x) \rightarrow (\exists y)(\Omega_1(y) \wedge \phi))$ or, in a more perspicuous abbreviated form, $(\forall x \in \Omega_0)(\exists y \in \Omega_1)\phi$, with quantifiers of greater depth bounded by Ωs with greater subscripts. The axioms of $\mathrm{Fin}(T)$ are all regular relativizations of axioms of T, axioms that state that $\Omega_p \subseteq \Omega_q$ when $p < q$, and axioms stating that the Ωs are order indiscernibles, that is, that if ϕ' and ϕ'' are both regular relativizations of ϕ, then every formula of the form $(\forall x_0 \in \Omega_p) \ldots (\forall x_n \in \Omega_q)(\phi' \leftrightarrow \phi'')$ that is the regular relativization of some formula is an axiom. Mycielski and Pawlikowski showed that every model

The indefinitely large is an epistemic, context-relative notion that itself involves no commitment to anything infinite. But that raises a question: on what grounds could we legitimately extrapolate, taking our experience of the indefinitely large to be genuinely relevant to what the set-theoretic infinite is like?

Any adequate justification whatsoever for our belief in the truth of present-day set theory and our consequent commitment to the existence of a proper class of entities must meet (at least) two desiderata: it must show why ordinary infinitary set theory—with its strong ontological commitments—is to be preferred to finite set theory, and it must show how knowledge of the infinitistic concepts of set theory could have arisen. Extrapolation can easily be seen to meet both desiderata, and it will therefore serve as a justification for set theory once it has been shown to be a motivated procedure and not just a technical trick.

Note that when I say justification of a theory, I mean justification in a very strong sense—justification, or in other words, adequate reason to believe, that the theory is true and the entities it mentions exist. That must be clearly distinguished from weak justifications that attempt merely to show that a theory is the simplest, or interesting and worthy of study, or admirable as a work of art. When I write about justification in the rest of this article I always mean strong justification. I shall occasionally write "strong justification" instead of just "justification," but that is only for emphasis—I mean the same thing in either case.

Here is how extrapolation can be motivated: When we apply mathematical theories to physics, if our intent is to model not our measurements but an objective world of physical quantities instead, then what we take to be in the available domains of finite mathematics must be what is in the physical world. Precisely because the physical world does not change with our epistemic situation, all the domains are set equal, and a finite theory is extrapolated into an infinitary one. The extrapolation is justified by the intended application, and its result is preferable to the underlying theory in finite mathematics to just the extent to which we value theories of the physical world over theories of our measurements.

The theory $\mathrm{Fin}(T)$ will always have better empirical support than T: the two explain and predict the same phenomena, but T has a much larger ontology with no explanatory gain. Adding the consideration of objectivity motivates the increased ontology, and T is the only theory with that ontology that preserves the explanatory success of $\mathrm{Fin}(T)$. On the account proposed here, our intuitions underlying a finite

---

of T has a substructure with finite domain that can be expanded to a model of T_f for every finite subset T_f of $\mathrm{Fin}(T)$. Thus, a theory $\mathrm{Fin}(T)$ has no commitment to the infinite, even when T does. Moreover, they showed that proofs in T of ϕ and in $\mathrm{Fin}(T)$ of a regular relativization of ϕ are closely related and of essentially the same complexity ([16], [10]: 273n). Thus, the reduction in ontological commitment comes for free, with no loss of expressive or proof-theoretic capacity. The Ωs are to be thought of, on my account, as bounds of current knowledge: the extension of Ω_0 is what is most immediately available to us (numbers actually written down, marbles in the room), while the extension to higher Ωs represents items that are more remote (sums of numbers actually written down, marbles in the next room). A theory $\mathrm{Fin}(T)$ collapses to (a notational variant of) T when axioms are added stating that all the Ωs are equal, which follows, for example, from $\Omega_1 \subseteq \Omega_0$, in virtue of the order indiscernibility. That is what I call extrapolation.

theory are extended to its extrapolated counterpart in a manner precisely analogous to that in which our observations yield a picture of the objective physical world, and the resulting extension of finite set theory is what gives rise to our concepts concerning the infinite.

What would constitute an argument that the proposed justification of set theory via extrapolation is the best argument for the legitimacy of set theory, that is, the best argument that set theory is true and that the entities to which it is committed are real?[2] To support the justification of set theory via extrapolation is in part to show that that method of strongly justifying our set-theoretic practices is superior to its rivals. That will be clear, provided the justification via extrapolation meets some minimal adequacy conditions, because no other strong justification for set theory that has been proposed both establishes the superiority of infinitary mathematics to finite mathematics and shows how the concepts behind infinitary mathematics could have arisen: Quine–Putnam indispensability [21, 22, 20], for example, or positions that require no more than the consistency of a theory to justify its use (for example, Balaguer's full-blooded Platonism [1] or Hilbert's program [8]) do not make it possible to distinguish between any infinitary theory and its counterpart within finite mathematics, since the two always yield parallel theorems, have equal consistency strength, and are notationally equally simple ([16] [10]: 273n).[3] Moreover, such positions make no attempt to explain how we could have come up with a theory of the infinite in the first place, and there are, as I have argued elsewhere ([10]: 2, 8, 140, 162, 164–165, 179, 246 247, 316–317), special problems in seeing how we could have come up with such a theory: we have no experience of any actually infinite systems or at least no experience that cannot be reasonably interpreted as experience of finite systems, as is shown, for example, by the fact that some physicists think that nature is finite, cf. [8].

Positions that attempt to introduce the infinite directly like the iterative conception [4, 5, 2, 17] or various attempts to employ the idealized capacities of an eternal mathematician (for example, by Cantor, see ([7]: 15, 35–36, 44), also [17, 28], ([9]: Chapter 6, 44)) presuppose the infinitary concepts it should be their job to explain. For that reason, they are inadequate justifications of infinitary mathematics.

Why do we need a justification for infinitary mathematics at all? We don't. As Shapiro has argued (for example, in the Preface and Chapter 2 of [25]), there is no good reason to expect or require any justification of any mathematical theory to whose truth we are committed. But that does not mean that it would not be desirable to have a justification, supposing that one can be obtained. In a philosophical spirit comparable to the mathematical spirit of reverse mathematics (see, for example, [26]), such justifications are desirable because they show what depends on what. Moreover, without a nontrivial justification for infinitary mathematics, it is hard to see what motivation there could be for adopting it in preference to finite mathematics.

[2] As I discuss below, nothing I say here should be construed as an argument that accepting set theory *requires* a prior argument for its legitimacy.

[3] One can, in doing finite mathematics, leave the bounds to indefinitely large sets implicit, in which case the notation and proof system become identical in surface form to those of ordinary mathematics.

I discuss the Quine–Putnam indispensability argument even though, as I have already noted, it cannot justify belief in infinitary set theory since it cannot distinguish infinitary set theory from its counterpart in finite mathematics. I do so because I take extrapolation to rely on a modified version of Quine–Putnam indispensability, and Quine–Putnam indispensability has come under attack on a couple of different fronts of late [12, 27, 13].

The Quine–Putnam indispensability argument is fundamentally a form of inference to the best explanation: it is the idea that if a mathematical theory is an indispensable part of our best theory of the physical world, that is, an indispensable part of the physical theory that best accounts for our observations and measurements, a theory that we take to be true with good reason, then that indispensability constitutes evidence for the truth of the mathematical theory. I rely on that idea to justify the use of theories of finite mathematics in accounting for observations and measurements, and so I must defend it from its other critics. My criticism of Quine–Putnam indispensability as it is ordinarily applied is that *not enough mathematics is indispensable,* and so Quine–Putnam indispensability doesn't do the work for which it has been employed. It can best be used to yield a commitment to theories of finite mathematics, which theories provide, by construction, the ability to explain all measurements and correlations between them explained by corresponding ordinary theories and to do so with far fewer ontological commitments.

The first attack on Quine–Putnam indispensability: Set theory is used in all physical theories, and so the success of any one physical theory is no evidence for or against set theory [13, 27]. That argument is wrong: Aristotelian physics, for example, presumably did not use set theory, and there may well be physical theories that make use of intuitionistic or category-theoretic mathematics.[4] If theories including a certain subtheory (like those including set theory) are so clearly superior in explanatory power to theories not including that subtheory that any theories not including it have dropped out of serious consideration, then that is the best possible evidence that the subtheory is a part of the best explanation of scientific phenomena.[5]

The second attack: Set theory is only used in idealized physical theories, and since we do not think the idealized theories true, they cannot license any commitment to the truth of the mathematics they employ [12]. But (a) idealized theories are in fact the best we have (and so a form of inference to the best explanation legitimately applies), and (b) there is every reason to believe that less-idealized theories—to the extent to which they become available—will be just as mathematical. That removes any clear distinction between those mathematical parts of an idealized theory that we take seriously and those that we don't and hence removes any possibility of making

[4] Mycielski was kind enough to point out to me that theories using intuitionistic mathematics or category theory are relevant here.

[5] The somewhat abstract phrasing in the text about subtheories is necessary because there are, in fact, perfectly good physical theories that do not use set theory—those based on finite mathematics—which provide explanations that are at least as good as those given by the theories that employ set theory. I therefore abstract from the occurrence of set theory in the attack. Set theory is now understood only as appearing in a particular instance of a more general argument.

use of such a distinction to dismiss the Quine–Putnam indispensability argument. For example, we don't take all aspects of continuous models of fluids seriously. But more realistic models start from the same equations, and so the same mathematics remains just as indispensable.

Now, let us return to the justification of extrapolation proper. I shall not concentrate on the justification via extrapolation of a theory of sets, but only the justification via extrapolation of elementary analysis. The justification I shall present of extrapolation from a theory within finite mathematics to analysis is intimately connected with the intended application of the real numbers to the measurement of physical quantities.

I believe that the strong justification via extrapolation of analysis to be discussed should be taken to extend to an analogous justification of set theory. Functions from ordered systems of real numbers to ordered systems of real numbers play a role in the measurement of physical quantities analogous to that of the real numbers—such functions are used in the measurement of physical fields—and it is such functions and the need to integrate partial differential equations that led to set theory ([10]: 39–41, 47, 49–51). I therefore see the situation with respect to the justification of set theory via extrapolation as being highly analogous to and intimately connected with that of the justification of analysis via extrapolation. Moreover, modern analysis, originating as it did with Weierstrass, Dedekind, and Cantor, is itself set-theoretic, and so the justification of elementary analysis via extrapolation is already a justification of a part of set theory via extrapolation. In addition, whether or not I am right that the justification of analysis via extrapolation directly extends to a justification of set theory via extrapolation, the success of applying extrapolation to analysis is itself evidence for the general validity of the procedure of extrapolation, to whatever theory (including set theory) it is applied.

The description of the real numbers as being used to give the values of the physical quantities in our physical theories, while accurate by our lights today, is not true to the history of the relationship between real numbers and physical quantities. Up until the "arithmetization of analysis" in the mid-19th Century, the real numbers and the physical quantities we now take them to represent were not distinct and separate things. The real numbers were taken to be geometric quantities, and the geometry was the interpreted geometry of actual physical space, and so, until comparatively recently in the development of our conceptualization of the real numbers, the real numbers were themselves actually taken to be physical quantities—ratios of lengths of lines in physical space and the like. See [10]: Chapter 2 for a synoptic history and references. Thus, our present-day conception of real numbers as being appropriate to reflect certain properties of physical quantities is not some desideratum of physicists or philosophers of science artificially imposed on the real numbers from the outside. It was and is a central part of the development and conceptualization of the real numbers that they must be like physical quantities in certain central respects. My argument relies on that necessary resemblance.

To say that the length of a rod is between 4.5 and 5 meters is just to say that nine meter sticks lined up is shorter than two aligned copies of the rod and that ten meter

sticks lined up is longer than two aligned copies of the rod. Moreover, given two (systems of aligned copies of) rods, there is a fact of the matter about whether they are the same length and, if they are not, there is a fact of the matter about which is longer. Analogous facts apply to all physical quantities, not just length—often for the simple reason that the measurement of many physical quantities is performed by first associating those quantities with suitable lengths, temperature in simple cases to lengths of columns of mercury, times to distances traveled by beams of light, and so forth.

Every physical theory takes as basic some system of physical quantities (which may not be lengths in all cases but other physical quantities—for example, relative statistical frequencies or propensities) for which such comparisons are possible. That is to say, according to every physical theory, it is possible, once a choice of unit[6] has been made, to compare any physical quantity with any rational multiple of the unit. Note that I am here discussing the presuppositions of physical theories, not the experimental or epistemological problems that are frequently involved in actually making the comparisons.

It is a presumption of contemporary physics that every physical quantity is already less than, greater than or equal to every rational multiple of a given unit of that quantity, whether or not a comparison has actually been performed. Note that that is a consequence of the assumption that our physical theories concern an objective world in which the outcomes of suitable physical operations—in particular, comparisons of physical quantities with rational multiples of a unit—are univocally determined quite independently of any human intervention or any human knowledge of the outcome.[7]

The assumption of the objective nature of the physical world has been disputed by idealists, phenomenologists, empiricists, instrumentalists, and so forth, and, arguably, by lattice gauge field theorists, but I shall just put such doubts aside here—they are part of a different and much larger topic. I believe that doubts about the objective nature of the physical world inevitably lead to derivative doubts about the status of infinitary mathematics. Infinitary mathematics cannot be strongly justified except on the basis of some assumption like that of the objective nature of the physical world.

I just argued that it is a fundamental assumption of contemporary physics that every physical quantity is less than, greater than, or equal to—comparable to—every rational multiple of a given unit of that quantity. That follows from the presupposition that a comparison can always be made between any physical quantity and a rational multiple of a suitable unit and the presumption that the results of such comparisons are fixed in advance of any actual act of comparison. That means that a physical quantity takes on a real number value, since the real numbers are, one might even say by definition,

[6]Anyone who is uncomfortable with the need for a choice of unit should feel free to replace all of my examples with ones concerning dimensionless constants.

[7]It is just as true of quantum mechanics as it is of more familiar theories that each quantity is comparable to every rational multiple of a suitable unit, when one restricts attention to suitable quantities—what is surprising about quantum mechanics is not that there are no such quantities but that certain familiar quantities are not of that type.

precisely the objects that realize every consistent system of comparisons with rational numbers. We can see that, for example, by noting that a quantity determines the set of all rational multiples of the unit less than it and the set greater than it: that is, it determines a Dedekind cut. Though I have argued here that every physical quantity takes on a real number value, that by no means entails that every real number is the value of some physical quantity.

The preceding considerations concerning the possibility of comparisons of lengths show that, insofar as real numbers are used to give the values of physical quantities in an objective world, the real numbers obey trichotomy (that is, each is less than, greater than, or equal to any other). Thus, our considerations have ruled out the intuitionistic reals, which do not obey trichotomy. The intuitionistic reals certainly form a coherent and mathematically interesting system. That system, however, is not the one of interest here, not the one that coheres with the desideratum that the real numbers be suitable for serving within mathematical physics as the values of objective physical quantities. I have argued for the necessity of the assumption of trichotomy for systems of real numbers suited for a particularly important and historically central application of the real numbers. I have not shown, nor do I think it could be shown, that there could not be reasons for adopting other systems for other purposes.

Let us explore how we represent (the values of) measurements as opposed to the physical quantities of which they are measurements. In arriving at the conclusion that physical quantities are represented by real numbers that obey trichotomy, we have made use of *every* rational multiple of a unit, an infinitary notion. In contrast, we conceive of measurements as having known results. Since all measurements available to us in a given experimental context are of some bounded accuracy, one natural and familiar model is that of numerical analysis: a finite grid or lattice of points all, for example, with coordinates of some fixed number of digits, with a corresponding theory given in terms of finite differences. That picture is overly simple in several respects. Measurements of different quantities may be of different accuracies, and measurements even of the same quantity may be of different accuracies in different ranges. Methods of measurement will in general only cover finite ranges.

There is no reason to suppose that the potential values of measurements form a nice, regular array. Instead, in general, the domain of potentially available values must be taken to be nothing more than a finite set of values of the appropriate sort, usually n-tuples of rational numbers. Moreover, in many contexts it is important to take account of the possibility of values of increased accuracy, that is, of the possibility of more refined measurements or of computed values of greater accuracy than the measured ones. After all, it is an all-too-familiar fact that the outcome of measurement plus modelling is often nothing more than an indication that more accurate measurements will be required—a fact that, among many other things, leads to a host of techniques in numerical analysis, including those involving adaptive approximation, which involve interactively moving from grids to other, finer ones. The moral is that not one but a sequence of sets of available values, representing improved possibilities of precision, each set including the previous one, is needed to model the process of measurement.

A theory of how measurements on a system behave should be independent of the details of current measurement technology, and so the theory should be such that it works for any suitable sequence of sets of available values—so that, for example, a present sequence can be extended as measurement technology improves, and so that the theory is not sensitive to the actual way in which technology improves—a changed history that skipped a step, leaping forward faster, or, similarly, a slowed down history that interpolated more steps will lead to different sequences of sets of available values, but that should not, of itself, necessitate any changes in the theory. For the reasons just given, a general theory of measurements on a physical system must be a theory on a finite increasing sequence of finite sets of available values such that the members of the sequence are order indiscernibles with respect to what is expressible in the theory—indiscernibles so that adding to or removing members from the sequence does not require changes in the theory. Thus, finite mathematics ([14, 15, 19], see also [10]: 268–285), with its hierarchy of finite domains each of which can serve as a finite system of discrete possible values for measurements, each member of the hierarchy encompassing more values and hence permitting greater resolution, that are order indiscernible, seems tailored to a theory of measurements that have the character of being available with finite, but, through change of the experimental situation, increasable accuracy.

Our theory of physical quantities, how they evolve and constrain each other, whatever it may be, is empirically answerable to a parallel theory of measurements, how their results evolve and constrain each other. That is just to say that we do not test hypothesized relationships between physical quantities directly, but only through tests of hypothesized relationships between measurements of physical quantities, which may be compared with the results, of finite accuracy, of measurements that have actually been made. Physical quantities themselves, far from having known values, are conceived of as having definite, context-independent values quite independent of any knowledge we may have acquired of them through measurement.

Consider, for example, some ordinary mathematical theory of a physical system, say, concerning position along a line. A counterpart theory to it in finite mathematics will be a theory concerning, not positions on a line, but finite approximations—measurements—of positions along the line. The intended model of the theory concerning positions on a line may naturally be taken to include a copy of the real numbers. The counterpart of the theory in finite mathematics will be such that every finite set of its axioms has a model that is a finite substructure of, a finite approximation to, the full intended model. As more axioms are added, they put more constraints on the models. Larger sets of axioms embody commitments to finer and finer approximations.

Every theorem of the ordinary theory has a finite analog that can be proved using some finite set of axioms of the finite counterpart theory. Thus, a finite analog of every theorem holds in every sufficiently fine sequence of approximations, and finite approximations are always available. In the finite theory every quantifier is bounded by a predicate representing an indefinitely large set, where indefinitely large means in

part that nothing is excluded from it[8] and that as additional constraints are imposed the indefinitely large sets are only constrained to increase, nothing is ever excluded.[9]

Any measurements, and any observed correlations between them, that can be accounted for by the ordinary theory can be accounted for in an exactly parallel way by a sufficiently large (though finite) subtheory of the counterpart theory in finite mathematics. Every finite set of facts expressible in the language of the ordinary theory, every finite set of inequalities representing measurements plus any finite system of equations correlating the values has a direct analog expressed in the vocabulary of the finite theory, and any part derivable from the ordinary theory will have its counterpart derivable in the finite theory. If the finite set of facts is consistent with the ordinary theory, then the counterpart in the vocabulary of the finite theory will have a finite model that is a substructure[10] of the intended model of the ordinary theory[11] that is a model of the formulation in finite mathematics of all the facts. Successive measurements and correlations can always be added on, imposing further constraints and thereby reducing the class of models.

The finite models of the finite theory reflect the measured and computed experimental evidence and theoretical correlates in a way that exactly parallels that of the ordinary theory, but the finite models only codify a commitment to a system of finitely many objects and correlations, tracking what our measurements and computations actually show.

As I have just described, the theories of finite mathematics account for our observations just as well as the more familiar theories that use ordinary mathematics, and they don't carry any commitment to anything infinitary, which is an advantage insofar as that means that they avoid a commitment for which we have no observational evidence. Moreover, the theories of finite mathematics are no harder to prove theorems in than the classical theories, as I have discussed above. So, what possible reason could we have for preferring ordinary theories to the theories of finite mathematics for applications in science?

What is in the world is just what there is, quite independent of our measurements of it, and it is context independent. After all, it is a fundamental part of objectivity that there are no varying degrees of being in the world analogous to the increasing degrees

[8] Speaking more precisely, "nothing is excluded from the set" means that for every finite subset T_f of Fin(T), every predicate Ω representing an indefinitely large set that is used in any of the axioms of T_f, for every model $\mathfrak{M}$ of the ordinary theory T, every finite model $\mathfrak{F}$ of T_f that is an expansion of a substructure of $\mathfrak{M}$, and every member m of the domain of $\mathfrak{M}$, there is a model of T_f that has $\mathfrak{F}$ as a substructure, that is an expansion of a substructure of $\mathfrak{M}$, and that has m in its domain and in the extension of Ω.

[9] Speaking more precisely, "they are only constrained to increase . . . " means that in any theory of finite mathematics and for any predicate Ω representing an indefinitely large set in that theory, there is no theorem of the form $(\forall x)(\phi(x) \rightarrow \neg\Omega(x))$. That is a consequence of the fact that nothing is excluded from any indefinitely large set.

[10] There is a slight simplification in the text: The vocabulary of the finite theory includes, in addition to the vocabulary of the ordinary theory, predicates for indefinitely large sets, and so the finite model is not literally a substructure as claimed: it is an expansion of such a structure to the enlarged vocabulary, an expansion with unchanged domain.

[11] "Intended" isn't playing any mathematical role here: the result stated in the text applies to any model.

of accuracy of measurements. Thus, the appropriate theory concerning positions along a line is one in which there is a single unvarying domain, not the hierarchy of increasing domains characteristic of measurements of increasing accuracy as modelled by the finite structures of finite mathematics. A theory of a fixed domain is a theory in which all the bounds on the quantifiers are fixed and equal to each other, and in our case it must be a theory that is parallel to our theory within finite mathematics concerning measurement along the line in order to provide parallel explanations of our observations and measurements. The theory concerning positions therefore must be the theory extrapolated from the one concerning the measurements, that is, the one obtained from that theory by moving from many domains to a single one by requiring that all the domains be equal. But that is extrapolation in the sense that I have proposed ([10]: 257–258)—it amounts to dropping the bounds to the sequence of sets, which are now superfluous. A theory extrapolated from any reasonable theory concerning measurements of position on a line will yield—without presuming any infinitary notions in advance—the result discussed above, that positions along the line are characterized by real number multiples of a unit ([10]: 273). Hence, extrapolation grounded in objectivity will justify the theory of the real numbers as the appropriate mathematical theory in which to formulate a theory concerning positions.

In like manner, a theory concerning measurements of values of a field in a three-dimensional space will, extrapolated, justify a theory of functions from the real numbers to the real numbers. Note that since a theory of measurements will allow arbitrary measured values for the initial conditions, with no functional dependence among them (that is almost a part of the definition of initial values), the extrapolated theory will have to include a theory of arbitrary functions, not a theory of some restricted class of analytic expressions, as its mathematical basis. But the theory of arbitrary functions and of the integration of partial differential equations just *is* set theory, which has therefore been justified on the basis of the metaphysical presupposition of the physical sciences that measurements are measurements of values of objective physical quantities in a world external to, and independent of, the measurements.

References

[1] Balaguer, Mark: 1998. *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.

[2] Boolos, Georg: 1971. The iterative conception of set. *Journal of Philosophy* 68: 215–231. Reprinted in [3], 486–502.

[3] Benacerraf, Paul and Hilary Putnam (eds.): 1983. *Philosophy of Mathematics*. Second edition. Cambridge, UK: Cambridge University Press.

[4] Gödel, Kurt: 1944. Russell's mathematical logic. In: P. A. Schilpp (eds), *The Philosophy of Bertrand Russell*, number 5 in Library of Living Philosophers, Evanston, Ill: Northwestern University Press, 123–153. Reprinted in [3], 447–469.

[5] Gödel, Kurt: 1947. What is Cantor's continuum problem? *American Mathematical Monthly* 54: 515–525. Errata, vol. 55, 151. Revised and expanded for the first (1964) edition [3], 470–485.

[6] Gödel, Kurt: 1986, 1990. *Collected Works*. Edited by S. Feferman *et al.*, New York: Oxford University Press.

[7] Hallett, Michael: 1984. *Cantorian Set Theory and Limitation of Size*. Oxford: Clarendon Press.

[8] Hilbert, David: 1926. Über das Unendliche. *Mathematische Annalen* 95: 161–190. Translation by S. Bauer-Mengelberg, pp. 369–392 [29]. Partial translation by Erna Putnam and Gerald J. Massey, 1964. Second edition of the Putnam and Massey translation, 183–201 [3].

[9] Kitcher, Philip: 1983. *The Nature of Mathematical Knowledge*. New York: Oxford University Press.

[10] Lavine, Shaughan: 1994. *Understanding the Infinite*. Cambridge, MA: Harvard University Press.

[11] Lavine, Shaughan: 1995. Finite mathematics. *Synthese* 103: 389–420.

[12] Maddy, Penelope: 1992. Indispensability and practice. *Journal of Philosophy* 89: 275–289.

[13] Musgrave, Alan: 1996. Arithmetical Platonism: Is Wright wrong or must Field yield? In: M. Fricke (ed.) *Essays in Honor of Bob Durrant*, Dunedin, New Zealand: Otago University Press, 90–110.

[14] Mycielski, Jan: 1980–1981. Finitistic real analysis. *Real Analysis Exchange* 6: 127–130.

[15] Mycielski, Jan: 1981. Analysis without actual infinity. *Journal of Symbolic Logic* 46: 625–633.

[16] Mycielski, Jan: 1986. Locally finite theories. *Journal of Symbolic Logic* 51: 59–62.

[17] Parsons, Charles: 1977. What is the iterative conception of set? In: R. E. Butts and J. Hintikka (eds.), *Logic, Foundations of Mathematics, and Computability Theory. Proceedings of the Fifth International Congress of Logic, Methodology, and the Philosophy of Science (London, Ontario, 1975)*, Dordrecht: Reidel. Reprinted [18], 503–529.

[18] Parsons, Charles: 1983. *Mathematics in Philosophy: Selected Essays*. Ithaca, NY: Cornell University Press.

[19] Pawlikowski, Janusz: 1989. Remark on locally finite theories. *Abstracts of Papers Presented to the American Mathematical Society* 10: 172. Abstract.

[20] Putnam, Hilary: 1979. *Mathematics, Matter and Method*, second edition. Cambridge, England: Cambridge University Press.

[21] Quine, Willard V.: 1948. On what there is. *Review of Metaphysics*, 2: 21–38. Reprinted with minor changes as pp. 1–19 of [23].

[22] Quine, Willard V.: 1960. Carnap and logical truth. *Synthese*: 12. Reprinted in [24].

[23] Quine, Willard V.: 1961. *From a Logical Point of View: Logico-Philosophical Essays*, second, revised edition. New York: Harper Torchbooks, Harper & Row.

[24] Quine, Willard V.: 1976. *Ways of Paradox and Other Essays*, revised and enlarged. Cambridge, MA: Harvard University Press.

[25] Shapiro, Stewart: 1991. *Foundations without Foundationalism*. Oxford: Clarendon Press.

[26] Simpson, Stephen G.: 1987. Subsystems of Z_2 and reverse mathematics. In: G. Takeuti (ed.), *Proof Theory*, Appendix. Amsterdam: North-Holland, 432–446.

[27] Sober, Elliot: 1993. Mathematics and indispensability. *Philosophical Review* 102: 35–57.

[28] Van Bendegem, Jean Paul: 1989. Foundations of mathematics or mathematical practice: Is one forced to choose? *Philosophica* 43: 197–213.

[29] van Heijenoort, Jean (ed.): 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press.

Department of Philosophy
University of Arizona
Tucson, AZ 85721-0027
USA

E-mail: shaughan@ns.arizona.edu

Russell's Absolutism vs. (?) Structuralism

Geoffrey Hellman

Abstract. Along with Frege, Russell maintained an absolutist stance regarding the subject matter of mathematics, revealed rather than imposed, or proposed, by logical analysis. The Fregean definition of cardinal number, for example, is viewed as (essentially) *correct*, not merely adequate for mathematics. And Dedekind's "structuralist" views come in for criticism in the *Principles*. But, on reflection, Russell also flirted with views very close to a (different) version of structuralism. Main varieties of modern structuralism and their challenges are reviewed, taking account of Russell's insights. Problems of absolutism plague some versions, and, interestingly, Russell's critique of Dedekind can be extended to one of them, *ante rem* structuralism. This leaves modal-structuralism and a category theoretic approach as remaining non-absolutist options. It is suggested that these should be combined.

1. Twists and Turns: Absolutism *and* Hints of Structuralism in Russell

Russell's book, *Introduction to Mathematical Philosophy* (1919), is almost as rich in philosophical views as it is in information for the general reader about foundations of mathematics. Early on, Russell writes,

> The question, 'What is number?', is one which has been often asked but has only been correctly answered in our own time. The answer was given by Frege in 1884, in his *Grundlagen der Arithmetik*. ([17]: 11)

Here we have a good expression of an absolutist stance: There is such a thing as the correct answer to the question, "What is number?", and, moreover, it is (essentially) the one Frege gave. (Cardinal) numbers are classes of equinumerous concepts (Frege), or—Russell would tolerate this much flexibility—equinumerous classes (Russell himself). Interestingly, this is immediately preceded by a brief discussion of an alternative "algebraic" or "structuralist" understanding of number concepts:

> It might be suggested that, instead of setting up '0', 'number', and 'successor' as terms of which we know the meaning …, we might let them stand for *any* three terms that verify Peano's five axioms. They will then no longer be terms which have a meaning that is definite though undefined: they will be 'variables', terms concerning which we make certain

> hypotheses, namely, those stated in the five axioms, but which are otherwise undetermined. … our theorems … will concern all sets of terms having certain properties. ([17]: 10)

But no sooner has he described the view than he proceeds to take it off the table. The passage continues,

> Such a procedure is not fallacious; indeed for certain purposes it represents a valuable generalization. But … in the first place, it does not enable us to know whether there are any sets of terms verifying Peano's axioms … . In the second place … we want our numbers to be such as can be used for counting common objects, and this requires that our numbers should have a *definite* meaning, not merely that they should have certain formal properties. (*ibid*.)

Nowadays, this must be regarded as a rather hasty dismissal. Even to Russell, one may justly claim, it should have appeared so. For, regarding the first reason, why should an algebraic or structuralist reading of number theory *per se* carry with it the assurance demanded, that of mathematical existence of a model? Of course, somewhere in an overall system one would want such assurance (to whatever extent that is possible), but neither *definitions* nor the use of grammatically proper names ever by themselves guarantee existence, as surely Russell knew, in connection, for example, with the ontological argument! Thus, Dedekind [6] first formulated the "Peano postulates" but gave them in the form of a definition of "*simply infinite system" ("progression"*, in Russell's terminology), but then attempted to prove existence separately. (The proof, as we know, went outside mathematics and was not rigorous, assuming "the totality of objects of Dedekind's thought", but the need for a separate proof was clear.) Moreover, Russell knew full well that one cannot fall back on the Frege–Russell definition of number to gain the assurance sought, for he had had his own bout with the Axiom of Infinity, to which he had had to yield by simply *assuming* it (as was done in Zermelo's set theory as well).

Answering the second objection requires scarcely more resources, surely not exceeding those at Russell's command. One can account for counting on an algebraico-structuralist reading as providing a bijection between enumerated objects and the relevant initial segment of *any* progression as indicated by the highest numeral reached. The Frege–Russell solution—in which the class enumerated *belongs to* the number—is elegant (in *this* respect) but hardly privileged. Set theory does it the other way, via bijections (even if one fixes on a particular definition of ordinals, say von Neumann's), and, to us, the elegance of the Frege–Russell account of counting appears as an artifact. This is especially so in light of the heavy price paid elsewhere in the system, e.g., reduplication of numbers at all type levels beyond their first appearance, not to mention their non-existence in the going set theory (ZFC).[1]

In fairness to Russell, he had already had raised a serious objection against Dedekind's structuralism (one we shall encounter again below in a modern context), and

an alternative version was not at hand. Perhaps it only took a mild lack of sympathy with the approach to tempt one into an easy dismissal.

In any case, Russell's "absolutism" was itself not absolute. A few pages after introducing Frege's great "discovery", we find the following clue that a touch of "imposition" is implicated:

> 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 … than to hunt for a problematical number 2 which must always remain elusive. ([17]: 18)

Moreover, as it turned out, the "correct definition" even had to be adjusted in light of paradoxes of naive class theory:

> But, for the reasons set forth [above, concerning paradoxes], if for no others, we cannot accept 'class' as a primitive idea … classes cannot be regarded as part of the ultimate furniture of the world. ([17]: 181–2)

As if this were not "taking back" enough, we are surprised to find in the end the following proposal regarding the nature of mathematical and logical propositions generally:

> We may thus lay down, as a necessary (though not sufficient) characteristic of logical or mathematical propositions, that they are to be such as can be obtained from a proposition containing no variables … by turning every constituent into a variable and asserting that the result is always true or sometimes true. … logic (or mathematics) is concerned only with *forms* … ([17]: 199)

Now, when we consider that *relations* (*or propositional functions*) as well as individuals count as constituents of propositions, then we realize that this criterion is met by construing mathematical propositions as formulated in higher-order logic without constants. Indeed, it suffices to work with second-order logic for number theory, analysis, and even set theory. Applying Russell's criterion to the case of number theory requires replacing '0' and 'successor' with an individual and two-place relation variable, respectively, in any sentence of second-order Peano Arithmetic (in which the other standard functions are explicitly definable). We may also replace the term

[1] To be sure, if one lowers one's sights and aims to recover just number theory and classical analysis, the advantages of the Frege–Russell definition can be realized in a demonstrably consistent second-order system invented by Boolos called "Frege Arithmetic" [3]. (The demonstration is relative to the consistency of second-order Peano Arithmetic, also called "classical analysis" in a formal sense.)

'number' with a unary relation variable, X, considered as the domain to which all quantifiers are relativized. *But this is just the proposal Russell considered and dismissed at the outset, as quoted above!* More precisely, Russell explicitly considered the replacement procedure applied to the Dedekind-Peano *axioms*. Applying the overall procedure to an arbitrary sentence A of the language of number theory naturally leads to the formation of a conditional of the form,

$$\forall R[\mathrm{PA}^2 \rightarrow A](\mathrm{S}/R),$$

in which 'PA^2' stands for the conjunction of the axioms and 'S/R' indicates systematic replacement of the successor constant with the relation variable 'R' throughout. (Here, to avoid clutter, we have dropped '0' as it can be introduced by definition from 'S'.) If A is logically implied by the axioms, the result is a truth of second-order logic. (If not, the result of replacing A with its negation yields such a truth, in light of the categoricity of the axioms, as Dedekind established.) Still better, taking the predicate 'number' itself into account as suggested, and generalizing, we obtain

$$\forall X \forall R[\mathrm{PA}^2 \rightarrow A]^X(\mathrm{S}/R),$$

where the superscript indicates relativization of all quantifiers to domain X. Again we have second-order logical truths under the same conditions.

By now we have come full circle (at least "up to negation!"), as this is already an expression of an eliminative structuralism applied to arithmetic, for it quite straightforwardly formalizes the claim that the truths of arithmetic are "what hold in any progression whatever." As Russell had phrased it, "our theorems … will concern all sets of terms having certain properties." Implicitly, indeed *almost explicitly,* Russell seems to have endorsed this view after all! Indeed, with but one more step—that of treating Russell's phrase, "and asserting that the result is always true", still more broadly to include possibilities of progressions, not just actual progressions—we arrive at the hypothetical component of a *modal-structuralist* interpretation, which simply prefixes the above with a necessity operator, as governed by a suitable modal logic (naturally chosen to be S-5, with certain restrictions, see [8]: Ch. 1, and [9]. The same procedure generalizes to analysis and many extensions, including set theories.

Furthermore, as Russell clearly recognized, the axiom of infinity, however it is stated precisely, while formulable in such a logical notation, is not a logical truth. Existence axioms or possibility axioms are still required. The possibility of a progression,

$$\Diamond \exists X \exists R[\mathrm{PA}^2]^X(\mathrm{S}/R),$$

must be assumed, or it can be derived from a still more elementary possibility assumption, which is not hard to formulate. That forms the "categorical component" of the interpretation, and it is needed to ward off the plague of "if-then'ism". (Of course, as is to be expected from strong medicines, there are side-effects.)

Clearly, Russell's absolutist stance was at odds with some of his actual proposals concerning the nature of mathematics. His "logicism" had structuralist elements

within it, and, as we shall see, his critique of Dedekind's views is of special relevance in assessing more recent efforts to articulate structuralism.

2. Varieties of Modern Structuralism and Their Challenges

Four main varieties of modern structuralism are readily identified: (1) set-theoretic ("STS"), based on model theory, (2) structures as *sui generis* universals (the "*ante rem*" structuralism of Shapiro [18] and Resnik [15]) ("SGS"), (3) modal-structuralism ("MS"), and (4) an approach based on category theory ("CTS"). These have been described in some detail elsewhere, see [10] and [11]. Here we recall briefly their leading characteristics.

STS goes back to the Bourbaki and today would appeal to model theory, with ZF as the background, as providing general concepts of mathematical structures as well as a theory of their interrelations and existence. Regarding number systems, although fixed set-theoretic interpretations are familiar (e.g., the finite von-Neumann ordinals for arithmetic, Dedekind cuts in the rationals for real analysis, etc.), these are seen as convenient ways to "fix ideas". Arithmetic truths are taken as truths in the language of arithmetic true in any (standard) model of PA (whether first or second order); similarly for the reals, the complexes, etc. With respect to these theories, STS is a version of *eliminative structuralism*, in that numbers as definite objects, referents of numerals and other singular terms, are eliminated in favor of multiple structures. This accords with the insight that the nature of the individuals is irrelevant; what matters are structural relationships, realizable in many isomorphic ways. More generally, all the spaces and structures familiar from the branches of ordinary mathematics are understood as set-theoretic structures, set domains together with functions and relations on them, whether or not the associated theories are categorical. Notably, however, set theory itself is not treated structurally: although one investigates various set-models of ZF and extensions, the ZF axioms themselves are not read "algebraically" or "structurally", i.e., as mere defining conditions on structures of interest; rather they are taken as *assertions* of truths outright, truths about "the cumulative hierarchy", itself "too big" to be a set.

SGS treats mathematical structures as universals, *patterns* in Resnik's terminology, answering to "what all particular systems (realizing key axioms) have in common", whether the systems are made up of concrete items or sets. One speaks literally of "*the* natural number structure", for example, made up of "positions" or "places" treated as abstract objects, not merely schematically as place-holders or "offices" to be filled by particulars. (Hence Shapiro's term, "*ante rem*", to contrast these structures from the "*in re*" systems, such as those of mereology (part-whole theory) or set theory.) It is such *positions* that are taken as the literal referents of numerals, etc., and the structural relations among them are taken as directly expressed by relation constants of our language, e.g., "successor", "addition", etc. Thus, this is a *non-eliminative* structuralism, closely akin to Dedekind's preferred conception. Axiom systems in

mathematics are at once defining conditions on structures *and* assertions about the ideal types they are taken to be about. On Shapiro's version, in the background is second-order logic and a list of (assertory) axioms directly governing the existence of *ante rem* structures; these resemble the axioms of second-order ZFC but with the addition of a Coherence Axiom, guaranteeing a structure answering to any "coherent" set of second-order conditions, where this is a new primitive corresponding to "realizable as a model" in the framework of STS, cf. [18]: 95.

MS has already been introduced above. Clearly, it is an eliminative structuralism, more so than STS in that it applies to set theories as well as those of ordinary mathematics. In its use of second-order logic, however, MS appears to require classes or Fregean concepts in the background. While such interpretations are possible, a class-free interpretation is available through a combination of mereology and plural quantification [9]. The objects of any entertained structures—now in the *in re* sense—are unspecified. So long as it makes sense to speak of wholes or combinations of them, they can occur in structures. Unlike SGS, no *special* "structural objects" are involved. Indeed, through the use of mereology and plurals, structures themselves as objects need not be recognized. One simply speaks of some, or these, or those, objects related in relevant ways among themselves or to other objects. A full theory of relations (and functions) is recovered, so that there need be no recognition of these as objects either. In this sense, MS can be understood as a (modal) *nominalistic* reconstruction. (However, it does not *have* to be taken that way. Its machinery is available to a wide variety of ontological frameworks, about which, officially, it can remain quite neutral.) Finally, it is an important feature of MS, in contrast to both STS and SGS, that it recognizes no absolutely maximal universe or domain for mathematics. It incorporates instead an *Extendability Principle,* that any domain whatever *can* be extended. Comprehension principles for wholes or for pluralities are restricted to be *extensional*: collection-like operations are confined, as it were, to within a world. (Officially, worlds or *possibilia* of course are not recognized; modal operators are primitive in the system, and are not explained literally as quantifiers.)

CTS is somewhat harder to describe because, although category theory and topos theory are well developed as branches of mathematics, a structuralist interpretation of mathematics in categorical terms remains somewhat inchoate. Sometimes it seems to be suggested that merely *formulating* mathematics categorically is enough to express a structuralist philosophy, since, indeed, CT has its own characteristic way of getting at mathematical structure, via morphisms between structures as point-like objects and via functorial relations among categories. That view is problematic, however, as it leaves unaddressed fundamental foundational issues, such as "What is the external, background logic?", "What existence axioms govern categories and topoi themselves?", "Are modal notions involved?", "Is an extendability principle recognized, as in MS, or should we take seriously an all-embracing category of categories?" (see [13]), etc. A full-fledged CT version of structuralism should address such questions. On one proposal, ordinary mathematics can be carried out relative to any number of topoi as universes of discourse, and this can be done without a set-theoretic background [1].

The result is a kind of *relativity* of ordinary mathematical concepts, and a distinction then arises between *invariant* mathematics (e.g., obeying intuitionistic logic, arising naturally inside a topos) and essentially *relative* mathematics, e.g., theory of classical continua for which special conditions on a topos are required (e.g., a choice principle, well-pointedness, etc.). Topoi are then viewed as "possible universes" for mathematics; there is no privileged, unique one. Is CTS then "eliminative"? With respect to number systems, surely it is; one may stipulate, for example, that a topos has "a natural number object", but this is not a unique structure in either Dedekind's or Frege's or Russell's sense. The case of set theory is less clear, as one speaks of "*the* category of sets", and uses boldface type to name it. How literally should this be taken? (Add this to the list of questions above.)

As will be evident already, none of these versions is problem-free. Vestiges of absolutism, present in STS and SGS, are not easily avoided. The goal of giving a structuralist interpretation of set theory (better theories) is sufficient motivation to look beyond STS. As will be brought out in the next section, however, SGS with its absolute "structural objects" is subject to a reinforced variant of Russell's critique of Dedekind's structuralism. Moreover, both STS and SGS suffer from commitment to a fixed, maximal universe for mathematics, violating the extendability principle, which is firmly rooted in mathematical thought and practice. Both these problems are overcome in MS and, pending further clarification, in CTS as well. But these approaches face their own characteristic difficulties. The various trade-offs are summarized in the following table.[2]

Main Varieties of Mathematical Structuralism and Problems

	STS	SGS	MS	CTS
Maximal universe	✓	✓	—	—
Sets exceptional despite multiple set theories	✓	—	—	—
Lack of equivalence types	✓	—	(ok)	(ok) (on relative interpretation)
Possibility of gross error (e.g., missing sets)	✓	—	—	—
"Positions as objects"	—	✓	—	—
Purely structural relations?	—	✓	—	—
Primitive modality	? (informal)	✓	✓	—
No account of math. existence or reference	—	—	—	✓

[2] In this table, a check-mark indicates that the problem in question does affect the version of structuralism, so that the aim of the game is to "draw a blank" (unless otherwise noted, as in the third row).

Concerning CTS, despite what we have said, we have taken our best guess as to how to fill in the boxes (trying to give the view the benefit of the doubt from a structuralist perspective).

The remainder of this paper will be devoted to explanation and discussion of some of the key boxes and relationships that emerge.[3]

Let us begin by explaining further the left-most column of problems. The first, "maximal universe" is clear, but it is worth quoting Mac Lane on the matter:

> Understanding Mathematical operations leads repeatedly to the formation of totalities: the collection of all prime numbers, the set of all points on an ellipse … the set of all subsets of a set … , or the category of all topological spaces. There are no upper limits; it is useful to consider the "universe" of all sets (as a class) or the category *Cat* of all small categories as well as CAT, the category of all big categories. This is the idea of a *totality* … . After each careful delimitation, bigger totalities appear. No set theory and no category theory can encompass them all–and they are needed to grasp what Mathematics does. ([14]: 390)

That STS violates this "open-endedness" is familiar from its commitment to a fixed "real world" of sets. But SGS also appears to violate it in it commitment to a totality of "all positions in structures" (at least, this is so on Shapiro's formulation in second-order logic). However, SGS improves on STS at the next row, as it does apply its "structural interpretation" to set theories generally, and it need recognize no maximal "totality of *all* sets". It automatically avoids questions such as, "Are the sets really well-founded?", "Do they satisfy Choice?", etc.; as long as the relevant axiom systems are coherent, there are structures answering to them, and none is ontologically privileged (although of course some may be of greater mathematical interest than others).

The third row, "lack of equivalence types", concerns the status of objects such as Frege–Russell numbers, or equivalence classes under isomorphisms preserving structure such as "ω-sequence", "countable, dense linear ordering without endpoints", "separable continuum", "complex plane", etc. Formation of equivalence classes is a natural way of passing from particular instances of a type of structure to "the type itself", but, as we know, this breaks down in set theory unless the instances are restricted to some level of the cumulative hierarchy (otherwise, proper classes are needed, raising problems of their own, especially from a structuralist perspective). SGS overcomes all this, as its abstract structures are supposed to be platonic archetypes, ideal exemplars answering directly to "what all instances have in common". MS, of course, recognizes no such animals, but counts this as a virtue, making sense of "shared structure" in more modest terms, e.g., by reference to satisfaction of the same axioms or via "structure preserving maps between structures", spelled out in the peculiar, MS fashion. CTS, at least on the "many topoi" relativist view suggested by Bell, also makes a virtue of the absence of absolute equivalence types. Again, shared structure is explained via external relations, i.e., morphisms and functorial relations, and one does not miss maximal or absolute archetypes.

The fourth row concerns impertinent questions occasioned by a literal, set-theoretical realism, such as "How do you know that the real-world power set of $\mathbb{N}$ is full?

[3] For a fuller discussion of the contents of this table, see my [10] and [11].

Maybe some subsets are missing!" Such questions do not arise on the other approaches. Each can tell its own story as to just how coherence of the notion of full power set is sufficient for mathematics (which leaves open the possibility also of denying coherence, which is a matter prior to structuralist interpretation).

The fifth and sixth rows bring us back to Russell.

3. Russellian Critique of Dedekind-Structuralism (with Help from Leibniz and Benacerraf)

As SGS illustrates, structuralism has sometimes been conceived in "absolutist" terms. The trick has been to posit special "structural", non-spatio-temporal objects, "pure places" in archetypal structures abstracted from all particular cases, whether themselves concrete or abstract. Shapiro calls the particular cases "systems", a more general term than "structure" in the *ante rem* sense. Structures qualify also as systems, but they can be self-exemplifying so that no "third man" regress is generated. (Since, however, plenty of systems are commonly thought of as already abstract, e.g., those built out of pure sets, or properties, etc., if recognized, I have called SGS a "hyperplatonist" view of mathematical structure.)

This procedure has its most distinguished antecedent in Dedekind ([5] and [6]), as is made clear from correspondence with Weber.[4] The natural numbers (or the reals, the topic of the Weber correspondence) form a unique, particular system over and above all other simply infinite systems (continua), exemplifying what they all share in common and supposedly lacking irrelevant features, i.e., those beyond distinctness from other objects and their roles as defined by structural relationships specified in the mathematics itself. (Tait calls the passage from particular realizations, e.g., set theoretic, to such pure archetypes "Dedekind abstraction". [19])

Now one may immediately question the idea of "purity". How can *any* objects fail to have mathematically irrelevant properties, such as arise from adventitious relations to external things, e.g., "being thought of by Dedekind", "being designated by the English monosyllable, 'nine' ", "being the number of planets", etc. Efforts have been made to salvage purity by distinguishing "essential" or "intrinsic" properties from the rest, but this is also problematic: surely on the conception being advanced, "being abstract", "being non-spatio-temporal", "lacking color or mass", etc., qualify as essential, but the mathematics is silent on such matters. In the *Principles,* Russell voiced misgivings along these lines:

[4] See [7]: vol. 3, 489–90. There is also, however, a letter to Lipschitz (dated June 10, 1876 [7]: §65, cf. [18]: 173) saying, "if one does not want to introduce new numbers, I have nothing against this …". While he may have preferred to think of (natural and real) numbers as special, "newly introduced" objects, Dedekind recognized that mathematics does not require this, and that what matters is "the holding of the right properties", which is compatible with an eliminative structuralism.

> It is impossible that the ordinals should be, as Dedekind suggests, nothing but the terms of such relations as constitute progressions. If they are to be anything at all, they must be intrinsically something; they must differ from other entities as points from instants, or colors from sounds … Dedekind does not show us what it is that all progressions have in common, nor give any reason for supposing it to be the ordinal numbers, except that all progressions obey the same laws as ordinals do, which would prove equally that *any* assigned progression is what all progressions have in common … .
>
> His demonstrations nowhere—not even when he comes to cardinals—involve any property distinguishing numbers from other progressions. ([16]: 249)

Nevertheless, *ante rem* structures are posited in SGS, in apparent defiance of these strictures. But perhaps it can be conceded that *places* in structures "must be intrinsically something", but that all that need amount to is that they just *are* different from the items Russell lists (and a lot more), that they depend essentially on intra-structural relations to other places, and that they are "grasped" by a kind of abstraction from "systems" as suggested. This is a stand-off, and further argument is needed to break it, one way or the other.

Two related objections have in fact emerged recently which challenge the very coherence of "*ante rem* structure" and "Dedekind-abstraction", in support of Russell. Both have a Leibnizian flavor, and the second comes with a topping of Benacerraf (retaining a similar flavor).

The first, due independently to Keränen [12] and Burgess [4], concerns the notion of *identity* of places in an *ante rem* structure. In a nutshell the argument goes like this: it seems that these "structural objects" should be individuated sufficiently by intra-structural relations (including functions) alone, without help "from outside" or from individual constants. (The reasons for this are given in detail in [12].) This yields a version of Leibniz' "identity of indiscernibles": any items bearing exactly the same intra-structural relations to other items must be not many but one. But this immediately implies that there can be no non-trivial automorphisms of the structure, i.e., that the structure is "*rigid*". Now, while some structures central to mathematics are rigid, such as the natural numbers, the reals (as a field), and segments of the cumulative hierarchy of sets, many are not. The complex field admits an automorphism interchanging i and $-i$; the additive group of integers admits an automorphism interchanging +1 and -1, etc. Indeed, such structures abound in mathematics, as Keränen observes: every group of order other than 1 or 2, any geometric figure with a reflectional symmetry, homogeneous Euclidean n-space (Burgess), etc. The rigid structures, though fundamental, are also rather special. Keränen explores possible responses the *ante rem* structuralist might give, including invoking *haecceities* (the property of "being just *this* thing and no other"), or giving no account of identity of structural objects at all, and finds them all unsatisfactory. Shapiro has replied that an account of identity is not really required. The debate over this continues.

Another response on behalf of SGS might be to appeal to *reduction* of all structures to rigid ones, perhaps technically possible if every structure can be adequately modelled set-theoretically. This seems alien to the whole SGS program and more in line with STS. In any case, the proposal is impotent against the second objection.

This objection, which I developed in [10], challenges the very intelligibility of *ante rem* structures altogether, rigid as well as non-rigid. Whereas the Keränen-Burgess objection takes structural relations as given and goes on to question the distinguishability of the structural objects as relata, this objection goes further in questioning the intelligibility of *purely structural relations*, in the context of putatively structural objects as relata. This is very much in the spirit of Russell's critique: if we do not appeal to relata as independently given somehow, but instead think of them as determined only through structural relations—as surely seems part and parcel of the *ante rem* structuralist view—then what have we to go on in specifying structural relations other than the axioms themselves as defining conditions? What, for example, can it mean to speak of "the ordering" of "*the* natural numbers", as objects of an *ante rem* structure, if, in turn we have no independent grasp of what these numbers are *apart from their position in that ordering*? As mathematical functions (treated extensionally in classical mathematics), the successor function of any given system or model is distinct from that of any other. So the archetypal one we're after isn't literally the same as any other we may be familiar with (such as "the next numeral" in our counting system, or "the next item" in some spatio-temporal sequence). Does it make sense to speak of "nextness itself", as a platonic abstraction from all instances, when anything whatever can be "next after" anything else in some system or other? It seems we can only make sense of "next" in the way that Dedekind himself did, namely, as *relative to a given function* φ, or arrangement of some sort, which is to say, in this case, relative to a given simply infinite system. Thus, we have a vicious circularity in the very notion of *ante rem* structure: such a structure is supposed to consist of *purely structural relations among purely structural objects, but understanding either of these depends on already understanding the other.* Given the objects, you can get the relations, and (if it weren't for the Keränen-Burgess objection) vice versa. But you can't get both at once!

One response to this (given by Shapiro in correspondence) is to deny that structural relations need be prior to places of an *ante rem* structure. For example, we have "finite cardinal structures" as degenerate cases of *ante rem* structures, in which there are a finite number of "places" but no relations at all (other than identity)! "The view must be that a structure is determined by its places and relations. Neither is prior to the other."

I agree that if we can somehow independently pick out the places, then we can speak of relations among them. This is a standard platonist procedure, falling back on, say, *designation* by certain singular terms in our language. But then the objects are not given "purely structurally", but by reference to external, and in this case, even contingent relations. But if we do not help ourselves to such extra-structural means of identifying the objects (which, I would suggest, is what we usually do in platonist mathematical talk), what sense does it make to speak of "places", if not *relative to given*

orderings or relations? With *ante rem* structures as proposed by SGS, we seem to be in the situation of having to succeed in referring to B (the *places)* as a precondition of referring to A (the *relations)*, *and vice versa*. But then you can't do both with neither. You have to do both with both, and that seems impossible.

What, then, of the "finite cardinal structures"? On reflection, these seem utterly baffling, indeed the ultimate offense against Leibnizian scruples. The "4 cardinal structure", for example, is supposed to consist of four distinct abstract things period. No structural relations at all are involved; it makes no sense to speak of one of them as "first" or as occupying any "position" in any intuitive sense. How is it that any is distinct from any other? Indeed, how can we make sense of referring to any one of them as opposed to any other, or mapping any one of them to or from anything else (which is essential if they are to exemplify cardinality)? Non-identity is not sufficient for this, and—unlike the natural numbers on an objects-platonist construe, according to which at least we have standard names for the objects, and unlike "identical bosons" of the particle physicist which at least "make a difference" by contributing to the total mass-energy, even if we cannot label them—nothing else is available. With these purported structures, as we shall see momentarily, the Keränen-Burgess and Hellman objections come together, in that every permutation of such objects would be a non-trivial automorphism ("trivially", so to speak, as there are no relations to preserve)!

This brings us to a final consideration. Suppose, for the sake of argument, that, despite the above argument, we *could* make sense of "introducing an *ante rem* structure", a rigid one, say, for the natural numbers. Call it $< \mathbb{N}, \varphi, 1 >$, where φ is the privileged successor relation and 1 the initial element of $\mathbb{N}$. Then we immediately see that indefinitely many other progressions, explicitly definable in terms of this one, qualify equally well as candidates to serve as the referents of our numerals, even if we require "freedom from irrelevant features", to whatever extent $< \mathbb{N}, \varphi, 1 >$ itself fulfills this. For we need only consider the result of permuting any (finite, say) number of items of $\mathbb{N}$ and adjusting φ and 1 accordingly. Any such structure, call it $< \mathbb{N}, \varphi', 1' >$, is able to serve as "the archetypal *ante rem* progression" every bit as well as $< \mathbb{N}, \varphi, 1 >$. Indeed, on what conceivable grounds can we call one but not any other "the result of Dedekind-abstraction". We cannot say, e.g., that "really 1, not $1'$", is "first" (if the permutation considered moves 1), for "first" makes no sense except relative to a successor function, and, relative to φ', $1'$, not 1, is first. The reader may well be reminded of Benacerraf's famous argument, that numbers cannot really *be* sets, since many progressions of sets are equally available to serve as natural numbers, and it would be absurd to say we are really speaking of one as opposed to any other [2]. Indeed, he generalized the argument to conclude that natural numbers cannot "really be" objects at all, and here, with *ante rem* structures, we have another example of why not. Hyperplatonist abstraction, far from transcending the problem, leads straight back to it.

This reminds us again of Russell's remark, quoted in the opening section above, on the "prudence" of settling for the class of couples rather than "hunt[ing] for a problematical number 2 which must always remain elusive". He was overly sanguine

about "the class of couples", but his worry about "a metaphysical entity about which we can never feel sure that it exists or that we have tracked it down" seems to us exactly on the mark.

4. Beyond Absolutism

The fundamental informal intuition behind structuralism is that, in mathematics, what matters are structural interrelationships and not the nature of individual objects. STS runs afoul of this in the case of set theory itself, normally taken as a theory about a fixed universe of absolute objects, "the sets". ("absolute" in the senses of "definite", "unconditional", and even "pure".) SGS, while even-handed in its application to set theories along with all other mathematical theories, nevertheless violates the intuition by positing special absolute objects, intended to stand in complete abstraction from any irrelevant, non-mathematical features, yet guaranteed by the mere coherence of suitable mathematical axioms implicitly characterizing them. Surely Russell was right to question the intelligibility of such "things" and to point to the futility of attempting to transcend the nature of individual objects by positing objects without any nature. And we have seen that deeper objections refining Russell's cut against the very notions of "*ante rem* structure" and "Dedekind abstraction". Moreover, as the table above makes clear, STS and SGS are "absolutist" in a second sense, in commitment to a maximal universe of mathematical objects, violating extendability, as already described. We are naturally led to seek non-absolute alternatives.

As the table also brings out, both MS and CTS bypass these problems of absolutism affecting either STS or SGS, although in very different ways. MS "abstracts" from the irrelevant features of objects of structures, not by positing special abstract objects, but by generalizing over *whatever* (suitably interrelated things) there might be, which can remain open-ended. Ordinary constants (e.g., numerals) *are* construed as "representing places in structures", not in the literal sense of SGS, but indirectly by indicating the relevant generalization: '2', in context, is paraphrased by reference to the third item of any progression (starting with '0'), and also as a convenient device in modal-existential instantiation on the assumption that progressions are possible in the first place (no pun here). CTS, on the other hand, avoids direct talk of items internal to mathematical structures entirely in favor of morphisms to and from other structures, these treated as "point-like", the "objects" of a category. In favorable circumstances (e.g., existence of terminals), suitable substitutes for "members of a structure" can be found, and, with ingenuity, CT paraphrases of ordinary set-theoretic conditions can be expressed. Like MS, CTS "transcends" the nature of individual objects (in the ordinary sense) not by postulation of special entities but by simply remaining silent about any such matters. MS, recall, enforces silence by sticking to mathematical conditions in its general statements concerning structures, and further by the thorough substitution of first- and second-order variables, as Russell ultimately suggested (cf. §1,

above). CTS, on the other hand, enforces such silence in its own clever way, by confining itself to properties got at via morphisms in a category and/or functorial relations among categories. Moreover, both MS and CTS avoid "absolutism" with regard to "the mathematical universe", again in their distinctive ways, MS by directly adopting principles of extendability and restricting comprehension to "within a world", CTS by not adopting any official ontology and proceeding informally in matters of (external) logic.

As the table also indicates, however, MS and CTS both confront problems of their own, in the case of MS problems relating to primitive modality, and, in the case of CTS, problems arising from failure to address fundamental questions concerning background logic, universes of discourse, and mathematical existence. If CTS avoids modality, that may be merely a manifestation of this stance. Interestingly, the "many topoi" view of Bell does address the question of universes of discourse without falling back on a domain of sets, and it hints at a modal formulation in which topoi are thought of as possible worlds for mathematics. Moreover, there are independent reasons for thinking that *some* appeal to modal notions is unavoidable, especially if one is to do justice to the non-contingency of mathematics and to its open-endedness, the extendability of its universes and its freedom to introduce ever new totalities, as described by Mac Lane above.

This suggests that we should seek a way of synthesizing MS and CTS, preserving their respective advantages while minimizing their liabilities. It turns out that this is indeed possible: the details cannot be presented here,[5] but the upshot is that the general, neutral apparatus of MS—mereology and plurals, without set-membership or classes—is sufficient to describe *large domains* (corresponding to set universes of inaccessible cardinality), and one can postulate these as mathematical possibilities, serving as sufficient backgrounds for categories and topoi, and for models of set theory as well. Of course, far, far less is required for ordinary mathematics, but this gives a unifying, structuralist framework for *extraordinary* mathematics, one in which topos theory and set theory can be developed side-by-side, neither taken as prior to the other. In this way, the respective structuralist insights of set theory and of category theory can be simultaneously preserved, without ever introducing absolute objects. The relativity of mathematics to choice of topos as suggested by Bell is incorporated in this synthesis, while at the same time the door is left open to classical mathematics as an objective enterprise.

References

[1] Bell, John L.: 1986. From absolute to local mathematics. *Synthese* 69: 409–426.

[2] Benacerraf, Paul: 1965. What numbers could not be. *Philosophical Review* 74: 47–73.

[3] Boolos, George: 1998. The consistency of Frege's *Foundations of Arithmetic*, 1987. In: G. Boolos, *Logic, Logic, and Logic*, Cambridge, MA: Harvard University Press, 183–201.

[5] For a fuller presentation, see [11].

[4] Burgess, John: 1999. Review of Stewart Shapiro (1997). *Notre Dame Journal of Formal Logic* 40: 283–291.

[5] Dedekind, Richard: 1872. *Stetigkeit und irrationale Zahlen*. Brunswick: Vieweg. Tr. as *Contintuity and Irrational Numbers*. In: W. W. Beman (ed.), *Essays on the Theory of Numbers*, New York: Dover, 1963, 1–27.

[6] Dedekind, Richard: 1888. *Was sind und was sollen die Zahlen?* Brunswick: Vieweg. Tr. as *The Nature and Meaning of Numbers*. Edited by W. W. Beman (ed.), New York: Dover, 1963, 31–115.

[7] Dedekind, Richard: 1932. *Gesammelte mathematische Werke 3*. Edited by R. Fricke *et al*. (eds.), Brunswick: Vieweg.

[8] Hellman, Geoffrey: 1989. *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Oxford: Oxford University Press.

[9] Hellman, Geoffrey: 1996. Structuralism without structures. *Philosophia Mathematica* (3) 4 : 100–123.

[10] Hellman, Geoffrey: 2001. Three varieties of mathematical structuralism. *Philosophia Mathematica*(3) 9: 184–211.

[11] Hellman, Geoffrey: 2003. Does category theory provide a framework for mathematical structuralism. *Philosophia Mathematica* (3) 11: 129–157.

[12] Keränen, Jukka: 2001. The identity problem for realist structuralism. *Philosophia Mathematica* (3) 9: 308–330.

[13] Lawvere, F. William: 1966. The category of categories as a foundation for mathematics. In: S. Eilenberg *et al*. (eds.), *Proceedings of the Conference on Categorical Algebra, La Jolla 1965*, Berlin: Springer, 1–20.

[14] Mac Lane, Saunders: 1986. *Mathematics: Form and Function*. New York: Springer.

[15] Resnik, Michael: 1997. *Mathematics as a Science of Patterns*. Oxford: Oxford University Press.

[16] Russell, Bertrand: 1903. *The Principles of Mathematics*. London: Allen & Unwin.

[17] Russell, Bertrand: first published 1919. *Introduction to Mathematical Philosophy*. New York: Simon and Shuster.

[18] Shapiro, Stewart: 1997. *Philosophy of Mathematics: Structure and Ontology*. New York: Oxford University Press.

[19] Tait, William W.: 1986. Truth and proof: The platonism of mathematics. *Synthese* 69: 341–370.

Department of Philosophy
Heller Hall 831
University of Minnesota
Minneapolis, MN 55455
USA

E-mail: hellm001@umn.edu

Mathematicians and Mathematical Objects

Robert S. D. Thomas

Abstract. The paper indicates why it is possible and even reasonable for mathematicians to be unconcerned with ontology of mathematical objects. Rather than such objects, mathematics is about 'relations between objects' or 'types of relation', in the words of Poincaré and Russell, the so-called objects being mere pronouns. No ontological position is taken up, and doubt is cast on the meaningfulness of deciding whether a pronoun exists as distinct from the existence of what the pronoun may be used to refer to in applied mathematics. Talking about what may or may not exist is made acceptable to philosophers, perhaps, by the pretence proposed by Mark Crimmins for semantics, based on that proposed by Kendall Walton for aesthetics.

I am grateful for this opportunity to raise with philosophers the puzzle of their different way of looking at mathematical objects. Speaking then as a mathematician to philosophers, I'd like to point out and to a small extent justify the view of some of us and make some small attempt to reconcile you to it. I shall have nothing to do with platonism on weekdays and formalism on the weekend. The attitude that I have myself and wish to justify somewhat and to reconcile you to is not caring about ontology, which despite the generality of its meaning seems always to be the study of things. Let me elaborate a little. It is my observation that mathematics deals, as Russell wrote in *The Principles of Mathematics,* with 'types of relations … the true subject-matter discussed, however a bad phraseology may disguise this fact' ([37]: §27, 23). While I do not share his apparent view that the usual way of speaking mathematically is just 'bad phraseology', it *does* disguise the subject-matter. It is bad only if one regards identifying the subject-matter as more important than saying something about it. Russell acknowledges hints of this perception in De Morgan, development of it by Peirce and Schröder, and a decisive debt to Moore for his view ([37]: §27, 24), something on which he agreed with Poincaré.[1] My way of describing what Russell calls 'disguise' is that, wishing to discuss relations, we postulate—in a peculiarly weak or thin mathematical sense of that word—objects to stand in those relations. If ontology were more often about relations, we should have some reason to take it more seriously.[2]

[1] 'Mathematicians do not study objects, but the relations between objects. To them it is a matter of indifference if these objects are replaced by others, provided that the relations do not change.'—page 20 of the English translation in [35]. The subject is kinds of n-ary relations, where $n \geq 2$.

[2] Frege distinguishes carefully between objects and relations and discusses relations in the part of the *Grundlagen* occupying §§70–83. Gray briefly reviews the nineteenth-century exploration of relations by Peirce, Schröder, and Frege in [16].

While philosophers might discuss fatherhood as they discuss justice and other abstractions, the mathematical way of discussing fatherhood would be to postulate a father and a child and discuss the relation in those terms, which have the function of standing for whoever is a father and whoever is his child. It may be that there is some sense in which such postulation calls the postulates into existence or some more attenuated form of life, but for purposes of the mathematical discussion that is a side issue. The point of this vagueness as it might be thought or disguise as Russell called it is that it considers systematically *whatever* does or might stand in the relations considered. The Zermelo and von Neumann integers are equally useful as examples so long as one considers the *relevant* relations that they exemplify and ignores the relations that they have but that are not under consideration. This is not wholly unlike the old-fashioned use of diagrams, which was not misleading as long as one did not get from them relations that were not actually postulated but happened to appear in the diagram. Because one is discussing *whatever*, the object terms of mathematics—or so I maintain—are pronominal not names.[3] As I have explained this at some length elsewhere [40, 41] and will say more about it later, I'm going to leave it now. The most extensive expression of the relational view, mainly expressed in terms of functions, is the book *Mathematics: Form and Function* by Saunders Mac Lane [31], universally ignored since it was published in 1986.

So I just go on to point out that I don't think mathematicians *should* pay much attention to ontology. I consider the existence question not sufficiently well defined to be amenable to discussion not at cross-purposes. This is my practical conclusion from reading something even as judicious as Charles Parsons's 1982 paper 'Objects and Logic'. On the one hand, one can find philosophers prepared to advocate the existence of any and all mathematical objects. On the other hand, one finds philosophers discouraging us from believing that any mathematical objects exist at all. I am not sure that they mean opposite things. The first sort of argument can have little effect because the sort of existence that the philosophers advocate is no use to mathematicians, and the second sort is better ignored because historically such negative arguments, whether coming from philosophers or mathematicians themselves, have only slowed the acceptance of mathematical advances like differential calculus and complex numbers. Another contrast to much the same purpose is that between the two social constructivist philosophies of Reuben Hersh and Paul Ernest, both trained in mathematics. In [20] Hersh says that mathematical objects are socially constructed and so exist. In [13] Ernest says that mathematics is socially constructed but its objects do not exist. The opposite ontological theses have no different consequences in the two versions of social constructivism. This much argument would, I think, suffice for a mathematical audience. I'd like before I'm done to make my case respectable to a philosophical audience.

I shall merely allude to Mark Balaguer's cleverly elaborate argument in [4] that there is 'no fact of the matter as to whether "there exist abstract objects" is true'

[3] In [16] Jeremy Gray mentions that Peirce called some of his variables 'pronouns' in 'On the algebra of logic', Amer. J. Math. 7 (1885), 180–202.

because it is not sufficiently clear what that says, '*i.e.*, because the sentence does not pick out a state of affairs that would need to obtain in order for it to be true' (personal communication, 2002 6 26). This is very close to my rejection of the narrower question about mathematical objects. I address only the narrower question.

I am curious about philosophers' reactions to an argument to the effect that we must be able to reason dependably about what does not exist as well as what does. This is an argument that ignores that mathematics *must* deal with the real and the unreal on an even footing—as Russell agreed when reviewing Meinong (Russell [36], quoted in Smith [38]: 308). It is based on what reasoning is *for*. The evolutionary advantage conferred on humans by our ability to reason is not that it allows us to think historically nor to think about the present, which rapidly becomes the past and so no longer real. It is our capacity to think about and therefore plan the future, much of which is not only unreal in the sense of not having happened yet (the hypothetical) but also unreal in the sense of being avoidable (the preventable). Preventable futures are among the most important and basic matters that we reason about.

One can say that mathematical practice is not or ought not to be influenced by ontology. I have heard this challenged with an equivocation on ontology, and so I want to make clear what I mean. Obviously mathematical practice has been influenced by ontology in the sense of what folks *think* there is; of course what we think influences what we do. What needs to be meant is that mathematical practice is not influenced by ontology in the sense of what there *is*; again of course we are influenced by being unable to talk sensibly along certain lines. So for the most part we do not talk about what involves us in contradiction, although that conclusion has lately been challenged.

In my reading of philosophy I have seldom encountered any argument to the effect that the existence question matters; it is just a major pre-occupation—just *there*, like Mount Everest. So I am not countering some major argument but rather trying to discourage the widespread implicit view that mathematical ontology is important—more important than considerations that would help us to understand mathematics better, *cf.* [32]. What has this matter to do with Russell? I think that Russell made a wrong turn when he took up what seems to have been Frege's prejudice that what matters is only what exists. From his initially favorable disposition toward Meinong, Russell turned against him and thereafter seems to have lost no opportunity to make fun of him [38]. By saying that Russell went astray, I do not mean to support Meinong against Russell. Meinong seems to have thought relations less existent than objects (merely what is often called 'subsistent'), while I think the reverse may be closer to the truth. Nor am I much interested in the locus of the Meinong-Russell debate, impossible objects. As soon as we determine that mathematical objects are impossible, we ignore them. The weakness of the mathematical sense of postulation is emphasized by the fact that we postulate them even for the sake of proving them impossible. But it is important that we can talk about them—so as to determine that they *are* impossible. What I mean about Russell's wrong turning is that, in dismissing the problems that Meinong raised or in thinking that he had given them their definitive solutions, he was mistaken. Perhaps the latter mistake should be attributed to others, but Russell

helped them by shifting from the respectful tone of his correspondence with Meinong to making fun of him later. To put this matter in my own terms, I need to point out that applied mathematics is different from and in some ways more difficult than pure mathematics. Meinong, as I understand what he was trying to do, was attempting to create a mathematical model of reasoning and implication—as was Russell. In all mathematical modelling efforts there is a practical decision how much to simplify the modelled in the model in order to have a tractable model. If one simplifies too much, one gets pure mathematics not usefully related to the area of application. If one simplifies too little, one has just a description of that area and no mathematics one can successfully work in. One might say that Meinong erred in the applied direction and Russell in the pure direction. That Russell appeared to win the day had the effect, among others, of slowing the effective modelling of this aspect of reasoning and implication. He prepared the desert landscape that Quine loved so much; the most fertile land is the temperate zone between Quine's desert and Meinong's jungle.

We all make mistakes, and in this turning I think Russell made a mistake, a mistake that the analytic tradition has persisted in up to the present day. Even Frege was not able to maintain the view with consistency. In his famous letter to Jourdain[4] he made up the first of the science-fiction examples beloved of analytic philosophers with his mountains Aphla and Ateb about which he reasons and obviously expects his reader to reason in spite of their referential emptiness.[5] This science-fiction stuff is the one place where I am fully in sympathy with the analytic rejection of discussing what is known not to exist. We clearly have a case here of 'do as I say, not as I do'. I have collected a little list of objectors to this prejudice [41].

The Quine-Putnam indispensability argument does not argue that the question is important, that our taking up a fixed view on this matter is indispensable, but rather assumes that science is the appropriate arbiter of what does and does not exist—a dubious premise in the light of science's helplessness in many intimate matters. And it concludes from scientific uses of mathematics that the so-called objects of mathematics must exist. This is just one more argument that mathematical objects do exist. My rejection of the sort of nineteenth-century scientism that this argument is based upon is of no interest in itself, but my counterproposal may be worth more. If

[4] *Philosophical and Mathematical Correspondence*, page 80, quoted in [14]: 14ff.

[5] Gareth Evans discusses the sort of make-believe that postulates things in section 10.2, 'Games of make-believe', calling them 'existentially creative'.

> An existentially creative game of make-believe can give rise to the possibility of someone make-believedly thinking of, or referring to, something without actually thinking of, or referring to, anything. ([14]: 360)

He makes it sound exceptional. In order to say carefully what he means he drops the common locution of the words' referring and falls back on the more correct reference by persons. He does mention the basis for its objectivity, mentioning the ability to 'focus on the same (empty) place in virtue of containing the same description' ([14]: 360ff.) and giving it some emphasis. The logical bog opened up by our habit of dealing competently with empty singular terms is alluded to by Evans in n. 43 on 37, but not its relevance to either make-believe (discussed much later in the unfinished book) or mathematics (which is not even in the index).

we are going to use science to study mathematical objects and the uses we make of them, then it should not be physics that we study but cognitive science, for instance the work of George Lakoff, Mark Johnson, and Rafael Núñez.[6] And the conclusions we can draw from such studies will not be primarily ontological. They may, however, be interesting or important. Having read the recent book by Lakoff and Núñez while I was preparing this paper, I can add that they will certainly be controversial.

I should like to see mathematical epistemology pursued in an ontology-neutral way. Arguments that mathematical or all abstract objects do not in some sense or other exist do not carry universal conviction. Arguments that these things do exist likewise fail to carry the day. Even if Balaguer is not right that there is no matter of fact on the existence question, there is quite definitely no *accepted* answer. It seems sensible therefore, to base at least some effort in other areas of philosophy of mathematics on that well-established empirical fact. Philosophy of science lives happily enough with the threat that any particular scientific entity may in a year or so dissolve into a number of other entities, may cease to exist as it was thought to do. Maybe real and maybe not is how Mario Bunge characterizes 'the scientific variety of realism' in [9]: 262: 'scientific theories refer to putatively real entities even though some such entities may turn out not to exist'. It is common for physical postulates to be thought not to exist, then to exist, then to be superceded. Let us do the same in mathematics and see what we can get. We might even get an explanation of why ontology (even regarded as about facts of the matter) is so unhelpful, what Azzouni calls the 'epistemic role puzzle' [1, 2]. At the *very* least, relations should get equal billing with objects, since the former are more basic and also more interesting. I have scandalized philosophers by my easy acceptance of this anontological attitude and my inability to express how it is coherent in language that they can accept. Kent Bach has a solution to the problem for fiction in his [3]: 216, but applicable more generally; reference is to the story (theory or whatever the context) in which the fictional object appears. This seems just a cop out—for mathematics if not for fiction. There is another view associated with Peter van Inwagen's [44][7] that Gareth Evans argues against in *Varieties of Reference*. It is the view that corresponding to unreal fictional characters in plays, for instance, there are real abstract objects, which are what authors create along with their fictional works and are what critics discuss. This would allow there to be no mathematical objects for the mathematicians but abstract simulacra for the philosophers. Evans points out that it ignores the need to engage in the make-believe for the sake of the background information, that the make-believe is *there*, and that the notion is overly sophisticated for the needs of such discourse ([14]: 367).

Mark Crimmins's application [11] of Kendall Walton's notion of make-believe [45] to Fregean modes of presentation is just the approach required.[8] I want to use this tool

[6] Lakoff and Johnson [28], Lakoff and Núñez [29].

[7] In [21]: 145 Harold Hodes mentions it as 'frequently attributed to Meinong' , and attributes it also to Kripke ('in a talk given at Cornell in 1982') and Searle (in 'The Logical status of fictional discourse' in *Expression and Meaning*, New York: Cambridge University Press, 1979), where I have been unable to find it.

to explain the irrelevance (in the sense I have explained) of ontology to mathematical practice. I am not attempting to enunciate a philosophy of mathematics but rather to explain why one can be interested in both mathematics and philosophy of mathematics and not be interested in having a definite ontology of mathematics.

Nothing here turns upon Crimmins's Frege, Russell, and Walton interpretations being correct; so when I say 'Frege' I mean Crimmins's Frege, whom I do not identify with the historical Frege. Likewise 'Russell' and 'Walton'. Crimmins interprets Frege as making non-trivial sense of identity statements like 'Hesperus is Phosphorus' by making them statements about the identity of the senses of 'Hesperus' and 'Phosphorus', non-trivial because of the different modes of presentation of the one object. And he interprets Russell as rejecting the attribution of relations in such cases.

> Rather, they assert complex claims about what *sorts* of things exist: for instance, 'Hesperus is Phosphorus' might assert that there exists a thing that is both the first star visible in the evening sky (in one season) and the last star visible in the morning sky (in another season). ([11]: 2)

Drawing attention to the phenomenology of such statements as 'Hesperus is Phosphorus' and 'Scott is the author of *Waverley*', Crimmins claims that we are involved in such cases in what he calls *semantic pretence*. We make believe that there are two things and then identify them. The make-believe is not permanent, but it explains our ways of thinking about non-trivial identities. He also points out that we can make meaningful statements about real persons in terms of fictional characters, using the example 'Ann is as clever as Holmes and more modest than Watson.' While this statement, like any statement on its subject, can be faulted for imprecision, its imprecision is not due to the fictional character of the fictional characters. Because such uses of fictional characters do not draw us into imaginative play as does reading a story, he calls this pretence *shallow*, but he insists that a degree of pretence is needed for such a statement to have any meaning at all. And I think he is right; the genuine truth condition of the Ann statement depends upon the fictional truth conditions of statements of the stories. He sets out the process of interpretation of the Ann statement in five steps:

> There is a transient context of make-believe, in which certain propositions are expressly made-believe …
>
> The fictional truth of certain other propositions is generated from reality [in this instance comparisons of Ann and the fictional characters] …
>
> It is fictionally true that I have expressed a proposition with my utterance of [the Ann statement] …
>
> Given the parameters of the make-believe, for it to be fictionally true that the proposition I have expressed with [the Ann statement] is true is for

[8] Stephen Yablo has another pretence-based explanation involving metaphor [48, 49]. J. P. Van Bendegem has also lately written on metaphor in this context [43]. How these writings relate to the so-called metaphors of Lakoff, Johnson, and Núñez is a large and perhaps difficult question.

> [a more complicated statement comparing the degrees of cleverness and modesty attributed to the characters and Ann's] to be true. …
>
> For my utterance to be genuinely, seriously true it is necessary and sufficient that it be fictionally true … ([11]: 6)

I have merely sketched Crimmins's proposal here. He is more elaborate and discusses literature that he thinks is relevant to his claims about what speaker attitudes are: are they pretending? Note that he claims a level of pretence that is shallow, just deep enough to get across the point that the make-believe reference makes.

> [I]n distinguishing thoughts by mode of presentation we talk as if we are distinguishing among the things thought about …we standardly talk as if we are referring to different things … ([11]: 14)

And of course he argues for his claims. He sums up.

> If we do quite standardly talk as if we are linguistically manipulating things when really we are manipulating modes of presentation, and if the pretense account gives a plausible story about how thing-language manages to do this work, then it matters not at all that we do not ordinarily think of ourselves as pretending that Hesperus and Phosphorus are two things. ([11]: 15)

What we *ordinarily* think does not matter; what we need is an explanation. Note, however, that in this explanation there is a shift from its mattering what there is to its mattering what we think. But what we think is based on objective evidence both of the Holmes stories and of Ann. And Crimmins claims to have saved the good part of Frege's and Russell's explanations without the parts he condemns as 'their worst problems' ([11]: 17).

> The pretense account offers a deep and phenomenologically plausible explanation for [standard] semantics. … an understanding of indirectly serious discourse that has been made possible by Walton's work in aesthetics. ([11]: 32)
>
> Walton suggests, somewhat tentatively, that in talking about what does and does not exist we might be engaging in a pretense to the effect that 'exists' expresses a discriminating property—a property that not everything has. ([11]: 33)

Such pretence is safer than the corresponding affirmation.

> This makes marvelous sense of a lot of puzzling talk about existence … ([11]: 33)

We can make non-paradoxical sense of talk about things that do not exist. We need no longer wonder with C. J. F. Williams [47] whether it is necessary to examine a

lot of blue buttercups before agreeing that 'none of them exist, that as a variety *blue buttercup* lacks existence'. We can also make sense of talk of what we do not know to exist. We treat them *as if* they exist even if we have paid no attention to Hans Vaihinger [42].

Between the extreme positions of what Mark Balaguer calls full-blooded platonism (held by James Robert Brown in [8] at least) and Hartry Field's fictionalism, which holds that all mathematical reference is empty, there are a large number of potential positions, many of them held by someone. Yet persons holding all those positions seem to use realist semantics and communicate successfully for mathematical and even some philosophical purposes. I do suggest that they are not communicating as well as would be convenient but not through failures of mathematical reference. Crimmins's pretence is what the fictionalist can with confidence be said to be doing, since mathematical abstractions are, for him, ontologically on a par with Sherlock Holmes, for whom Crimmins's picture was developed. Only the full-blooded platonist needs no pretence. All of the persons with intermediate positions and who differ on which mathematical objects exist or on what it means that those exist that they think exist,[9] can be understood to be operating under Crimmins's pretence when they appear to refer to what they do not affirm the existence of. We need to keep in mind that a statement to the effect that x is a topic of our thoughts or conversation is more a statement about us than about x, especially when there is no x. It makes no difference where the existence-non-existence boundary is, since the coping strategy is invisible; at an extreme the boundaries could be at different places for several interlocutors and no one need know or care where the boundaries are. Crimmins's picture looks exactly like realist semantics, being parasitic on it, and so it can be attributed to persons most of whom have never heard of it. It is intended, by me at least, as an explanation for philosophers of a phenomenon that they can observe for themselves, the mathematical level of postulation that does not involve any affirmation of existence, and which, for instance, Brown ([8]: 100) attributes to Hilbert.

The existence of mathematical objects, which seems to have mattered to the ancient Greeks, can accordingly be thought of as like training wheels. Needed at first, but as they are needed less and less whether they are there or not becomes of less importance, eventually being of no importance. This is not to say that training wheels and mathematical objects do not exist; it is just meant to be a vivid way of pointing out that to a bicycle rider and a mathematician, respectively, their presence and existence is of no importance. It is worth emphasizing that I am not saying that mathematical

[9] Interpreting the definition of the empty class in *Principia Mathematica* as meaning that 'the empty class is the domain of the logically contradictory' ([33]: 416ff.), Albert Menne concludes that the kind of existence of an object there is 'the denial that it is logically contradictory' ([33]: 417). He objects to a single sort of existence for 'Cologne Cathedral, a prime number between 35 and 40, a class that equals the complement of its complement, the wife of Zeus, and Gulliver' ([33]: 417), mentioning in passing that to take only the building as real would render the whole of mathematics fictitious, and claims that existence depends on the 'universe of discourse' we are taking as basic, whether 'real, mathematical, logical, mythological, [or] poetical' respectively. Menne proposes that all existential statements be taken relatively to a material predicate, allowing those he listed and such others as are needed.

objects do not exist; that is a question on which I have no reason to hold any fixed view. I do not see how others, even with attitudes I sympathize with, can be confident of their conclusions, *e.g.*, Azzouni, '… there is really no sense in which the objects postulated exist apart from their role as referents of particular grammatical structures (e.g., noun phrases) in systems' ([1]: 88). I am saying that it does not matter and am trying to explain, with the aid of pretence of arbitrary depth (to use Crimmins's metaphor), why it does not matter. To hold that all mathematicians *pretend* that mathematical objects exist would be factually wrong. What *is* a fact and needs explaining is why many don't care and so are free of ontological commitments that many philosophers demand (of them!)[10] that they make.[11] To return to Russell, I have been trying to point out why it is not unreasonable to maintain what Elizabeth Eames [12] says was his consistent post-Meinong position, that 'terms of logic and mathematics … may be

[10] Field makes such a demand in his second book [15].

> … a mathematical theory, taken at face value, is a theory that is primarily about some postulated realm of mathematical entities: numbers, or functions, or sets or whatever (or some combination, like numbers and sets together). You can't consistently believe the theory without believing in the entities it postulates. ([15]: 2)

Field is presumably using 'postulate' in the stronger philosophical sense. This quotation illustrates that the self-declared fictionalist Field ignores fiction, and it also suggests that mathematicians 'believe in' so-called mathematical objects. With the invention of non-euclidean geometries, mathematicians and others began not to believe in even the theories—as everyone had before—much less in the objects they are populated by. That is the relevance of the view that these stories, which are written like history, might be fiction; that we cannot know which, and that fortunately it does not matter. We know that the axiom sets for elliptic, euclidean, and hyperbolic geometries contradict one another. Unlike Field, Balaguer [4] does not ignore fiction, but his book does not helpfully relate mathematics and fiction.

[11] The problem of the objectivity of mathematics is one that can be thought to be intensified without ontology, in particular without real objects for our language to refer to. On the other hand, real objects to which the language is bound can be thought to make one's or everyone's reasoning irrelevant or at least to play second fiddle to the previously existing relationships to which one does not have access. A lack of pre-existing objects or ignoring them if they should happen to exist allows us the freedom to use various kinds of logic instead of having to depend on the exactly right one, whichever it is. Often referred to in this connection is Georg Kreisel, in particular his review of Wittgenstein's *Remarks on the Foundations of Mathematics* [25] and the paper [26]. I subscribe to 'the "old fashioned" idea', as expressed in that paper,

> that one obtains rules and definitions by analyzing intuitive notions and putting down their properties … What the "old fashioned" idea assumes is quite simply that the intuitive notions are *significant*, be it in the external world or in thought (and a *precise* formulation of what is significant in a subject is the result, not a starting point[,] of research into that subject).

But I think that the significant notions are relational, at bottom as objective and sharable as objects. Kreisel continues:

> Informal rigour wants (i) to make this analysis as precise as possible … in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them, and (ii) to extend their analysis, in particular, not to leave undecided questions which can be decided by full use of evident properties of these intuitive notions. ([26]: 138ff.)

In a less readily accessible paper Kreisel writes 'We consider the *objectivity* of certain notions and try to decide some questions about it without having to answer whether, in addition, some reality or realizability external to ourselves is involved.' ([27]: 20). The touchstones of objectivity in child and adult make-believe are props according to Walton, and props have their analogues for the world of Holmes in the texts and for mathematics in what formalism focuses on.

treated as entirely linguistic' ([12]: 125), with 'no implication of the non-existence' (*ibid.*) or existence intended.[12] Fiction and mathematics are slightly similar ways of using postulated objects to discuss relations. They and their offspring natural science represent the most successful way of dealing with relations.

Let me return now to the relational view of mathematical subject-matter. It looks a bit like a standard straw man called postulationism, and it may be useful to consider how it looks like postulationism and how they are different. A textbook definition of postulationism is that

> the axioms of a mathematical theory consitutes [*sic*] a set of specifications which define a structure that the mathematician wishes to study. ([30]: 19)

I call this a straw man because within two sentences the author, Hugh Lehman, shows us with Benacerraf's puzzle that we identify the real numbers as different sets simultaneously and so have a contradiction. This illustrates the wisdom of keeping our discussion at the pronominal level of whatevers and not sinking down to the substantive special cases; that is where the contradictions lie, and they are also not what we are wanting to talk about. Even categorical postulates allow for distinct interpretations. Interpreted numbers can be distances or masses, but not distances and masses. The axioms of a mathematical theory set out the relations holding among whatever satisfies this particular theory; they are statements about whatevers. When one says whoever is a bachelor is male and unmarried, one is saying something about human whoevers, but it is just absurd to ask whether the whoever exists or which whoever you refer to.[13] That is not the sort of thing such antecedent-free pronouns are for. This incompleteness of mathematical statements is, I assume, why Quine thought that unapplied mathematics is dubious in comparison with the applicable. The relational view looks like postulationism in Russell's *Principles*. That work famously begins by saying that 'mathematics is the class of all propositions of the form "p implies q," where p and q are propositions containing one or more variables, the same in the two propositions, and neither p nor q contains any constants except logical constants.' I found some difficulty in reconciling this statement with the more obviously abstractive attitude indicated in the context of geometry and dynamics in the preface,

> In pure mathematics, actual objects in the world of existence will never be in question, but only hypothetical objects having those general properties upon which depends whatever deduction is being considered; and these general properties will always be expressible in terms of the fundamental concepts which I have called logical constants. ([37]: xvii)

But the two are connected in §8:

> So long as any term in our proposition can be turned into a variable, our proposition can be generalized; and so long as this is possible, it is the busi-

[12] This attitude is exemplified by I. Jané [24], citing the dependability of proofs by contradiction where pretence is explicit.

[13] This is a comparison of pronouns with pronouns not of mathematical statements and analytic statements.

> ness of mathematics to do it. If there are several chains of deduction which differ only as to the meaning of the symbols, so that propositions symbolically identical become capable of several interpretations, the proper course, mathematically, is to form the class of meanings which may attach to the symbols, and to assert that the formula in question follows from the hypothesis that the symbols belong to the class in question. In this way, symbols which stood for constants become transformed into variables, and new constants are substituted, consisting of classes to which the old constants belong. ([37]: 7)

The book's initial sentence is not so much a recipe for producing mathematical statements as a description of the result of making statements that are already available into mathematical ones. As a recipe, it looks as though one can make anything up and call it mathematics, but as a description it is just a bit sloppy and applies to stuff that no one would call mathematics through a vagueness that is admitted by Russell in the 1938 introduction to the second edition.

In the light of discussion after this paper was delivered, it is perhaps more important to distinguish the relational view from structuralism. It is legitimate and indeed necessary and important to reify relations in mathematical work. Examples are ancient proportionality, which is a relation between relations, and the function-space concept. We would express proportionality by equating two ratios, which are reified comparison relations. We now perform algebra with functions as things, but until they had been reified they were themselves relations between independent and dependent variables. Of course they still are, as ratios are still comparisons. Relations do not stop being relations when they are reified, but when reified they can enter into other relations; it is a syntactic necessity. Structuralism, on the other hand, creates new highly problematic *objects* consisting of objects and the relations among them—or, as I should prefer to say, relations and their objects—and calls them structures. This is a philosophical rather than mathematical reification. It too is perfectly legitimate, but we do not yet know whether it is necessary and important. It does have the effect of submerging the relations and making them invisible instead of just ignored.[14]

Taking the relations as basic greatly reduces the importance of the axioms; they are a technique of accomplishing the organization we need. Postulation for me is just an agreement that this is what we're going to talk about today, what letters we're going to use to represent the 'things' standing in the relations we're going to discuss. It is the relations that are important, and they are usually relations that 'exist' in whatever sense relations can be said to exist, membership, inclusion, collinearity, non-collinearity, primality, *etc*. They are relations we know about and want to study scientifically.

[14] I acknowledge the justice of the complaint that I ought to give credit where credit is due to those approaches to mathematics that do as I wish. Unfortunately, any adequate evaluation of conceptual realism, attributed to Gödel (*e.g*., Wang in ([46]: §8.5.20) and actually defended by Daniel Isaacson in [23]), modal approaches (in [10], [17, 18, 19] and [39]), and approaches related to category theory as exemplified by John L. Bell's [7] (and his [5, 6] on which it is based) would require more space than this whole document and involve work of a nature and quantity beyond me.

References

[1] Azzouni, Jody: 1994. *Metaphysical Myths, Mathematical Practice*. New York: Cambridge University Press.

[2] Azzouni, Jody: 2000. Stipulation, logic, and ontological independence. *Philosophia Mathematica (3)* 8: 225–243.

[3] Bach, Kent: 1987. *Thought and Reference*. Oxford: Clarendon Press.

[4] Balaguer, Mark: 1998. *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.

[5] Bell, John L.: 1981. Category theory and the foundations of mathematics. *Brit. J. Philos. Sci*. 32: 349–358.

[6] Bell, John L.: 1986. From absolute to local mathematics. *Synthese* 69: 409–426.

[7] Bell, John L.: 1988. *Toposes and Local Set Theories: An Introduction*. Oxford: Clarendon Press.

[8] Brown, James R.: 1999. *Philosophy of Mathematics*. London: Routledge.

[9] Bunge, Mario: 1983. *Treatise on Basic Philosophy, Vol. 6*. Dordrecht: Reidel, 1974–1985.

[10] Chihara, Charles: 1990. *Constructibility and Mathematical Existence*. Oxford: Oxford University Press.

[11] Crimmins, Mark: 1998. Hesperus and phosphorus: sense, pretense, and reference. *Philos. Rev.* 107: 1–47.

[12] Eames, Elizabeth R.: 1999. Russell on "What there is". In [22]: 115–127.

[13] Ernest, Paul: 1998. *Social Constructivism as a Philosophy of Mathematics*. Albany, NY: SUNY Press.

[14] Evans, Gareth: 1982. *Varieties of Reference*. Edited by J. McDowell. Oxford: Clarendon Press.

[15] Field, Hartry: 1989. *Realism, Mathematics and Modality*. Oxford: Blackwell.

[16] Gray, Jeremy: 2001. Symbols and suggestions: Communication of mathematics in print. *Math. Intelligencer* 23 (2): 59–64.

[17] Hellman, Geoffrey: 1989. *Mathematics without Numbers*. Oxford: Oxford University Press.

[18] Hellman, Geoffrey: 1996. Structuralism without structures. *Philosophia Mathematica (3)* 4: 100–123.

[19] Hellman, Geoffrey: 2001. Three varieties of mathematical structuralism. *Philosophia Mathematica (3)* 9: 184–211.

[20] Hersh, Reuben: 1997. *What is Mathematics, Really?* New York: Oxford University Press.

[21] Hodes, Harold T.: 1984. Logicism and the ontological commitments of arithmetic. *J. Philos.* 81: 123–149.

[22] Irvine, Andrew D. (ed.): 1999. *Bertrand Russell: Critical Assessments. Vol. III: Language, Knowledge and the World*. London: Routledge.

[23] Isaacson, Daniel: 1994. Mathematical intuition and objectivity. In: A. George (ed.), *Mathematics and Mind*, New York: Oxford University Press, 118–140.

[24] Jané, Ignacio: 2001. Reflections on Skolem's relativity of set-theoretical concepts. *Philosophia Mathematica (3)* 9: 129–153.

[25] Kreisel, Georg: 1958. Review of Wittgenstein's *Remarks on the Foundations of Mathematics*. *British J. Philos. Sci.* 9: 135–158.

[26] Kreisel, Georg: 1967. Informal rigour and completeness proofs. In: I. Lakatos (ed.), *Problems in the Philosophy of Mathematics*, Amsterdam: North-Holland, 138–171.

[27] Kreisel, Georg: 1970. The formalist-positivist doctrine of mathematical precision in the light of experience. *L'Age de la science* 3: 17–46.

[28] Lakoff, George, and Mark Johnson: 1999. *Philosophy in the Flesh*. New York: Basic Books.

[29] Lakoff, George, and Rafael E. Núñez: 2000. *Where Mathematics Comes from*. New York: Basic Books.

[30] Lehman, Hugh: 1979. *Introduction to the Philosophy of Mathematics*. Oxford: Blackwell.

[31] Mac Lane, Saunders: 1986. *Mathematics: Form and Function*. New York: Springer.

[32] Maddy, Penelope: 1992. Indispensability and practice. *J. Philos.* 89: 275–289.

[33] Menne, Albert: 1982. Concerning the logical analysis of "existence". *The Monist* 65: 415–419.

[34] Parsons, Charles: 1982. Objects and logic. *The Monist* 65: 491–516.

[35] Poincaré, Henri: 1902. La grandeur mathématique et l'expérience. In: *La Science et l'Hypothèse*. Paris: Flammarion. English translation W. J. G.: 1952. Mathematical magnitude and experiment. In: *Science and Hypothesis,* New York: Dover, 17–34.

[36] Russell, Bertrand A. W.: 1905. Review of Meinong's *Über Gegenstandstheorie*. *Mind* 14. Reprinted in: D. Lackey (ed.), *Essays in Analysis*, London: Allen & Unwin.

[37] Russell, Bertrand A. W.: 1937. *Principles of Mathematics*. Second edition with new introduction. London: Allen & Unwin. First edition 1903.

[38] Smith, Janet Farrell: 1985. The Russell-Meinong debate. *Philos. Phenomen. Res.* 45: 305–350.

[39] Tharp, Leslie: 1989. Myth and mathematics: A conceptualistic philosophy of mathematics I. *Synthese*. 81: 167–201.

[40] Thomas, Robert S. D.: 2000. Mathematics and fiction I: Identification. *Logique et Analyse*. 43: 301–340.

[41] Thomas, Robert S. D.: 2002. Mathematics and fiction II: Analogy. *Logique et Analyse*. 45: to appear.

[42] Vaihinger, Hans: 1924. *The Philosophy of 'As If'*. Abridged translation by C. K. Ogden. London: Kegan Paul, Trench, Trubner.

[43] Van Bendegem, Jean Paul: 2000. Analogy and metaphor as essential tools for the working mathematician. In: F. Hallyn (ed.), *Metaphor and Analogy in the Sciences*, Dordrecht: Kluwer, 105–123.

[44] Van Inwagen, Peter: 1977. Creatures of fiction. *Amer. Philos. Quart.* 14: 299–308.

[45] Walton, Kendall L.: 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, Mass.: Harvard University Press.

[46] Wang Hao: 1996. *A Logical Journey: From Gödel to Philosophy.* Cambridge, Mass.: MIT Press.

[47] Williams, C. J. F.: 1992. *Being, Identity, and Truth.* Oxford: Oxford University Press.

[48] Yablo, Stephen: 1998. Does ontology rest on a mistake? *Aristotelian Society Supplementary Volume* 72: 1–33.

[49] Yablo, Stephen: 2001. Apriority and existence. In: P. Boghossian and C. Peacocke (eds.), *New Essays on the Apriori*, Oxford: Oxford University Press, 197–228.

St John's College and Department of Mathematics
University of Manitoba
Winnipeg, Manitoba R3T 2N2
Canada

E-mail: thomas@cc.UManitoba.CA

Russell's Paradox and Our Conception of Properties, or: Why Semantics Is no Proper Guide to the Nature of Properties

Holger Sturm

Abstract. In this paper I try to defend the following three theses: First, semantics is no proper guide to the nature of properties. Second, the phenomenon of self-instantiation is not essential to the nature of properties. Third, finding a solution to Russell's paradox is not the defining task of a theory of properties, in contrast to what most formal workers in the area of property theory seem to assume.

1. Introduction

Most formal accounts of properties are centered around some ambitious solution to Russell's paradox. This reflects the fact that the phenomenon of self-application or self-instantiation is considered to be essential for the nature of properties. According to this view, it is characteristic for the realm of properties that it accommodates entities that instantiate themselves, like the property of being a property, or the property of being an abstract entity. This, together with their ultra-fine identity conditions, is supposed to distinguish properties from all other kinds of entities.

Such a view usually rests on a semantical conception of properties. Properties are first of all regarded as entities that contribute to the meaning of language. More precisely, they are identified with the meanings or senses of certain kinds of linguistic expressions, such as predicates or their nominalizations. In the last three decades, this idea has also shown strong influence on the research within the field of theoretical or referential semantics. A number of philosophers and linguists have made use of properties as a foundation for natural language semantics, see for instance, [5, 7, 14].

The semantic conception of properties has the consequence that meaning has to be regarded as the main source of our understanding of what properties are like. To know the nature of properties is then, at least to a great extent, the same as to know how they contribute to the meaning of language. From the perspective of theoretical semantics, this means that to know the nature of properties is to know what theoretical roles they

perform, and how they manage to do so. In the extreme case, this leads to a position according to which properties are nothing but a particular kind of abstract or formal entities that have exactly the features that are required for making them suitable for the purposes of formal theories of meaning.

The main task of this paper is to call the picture just described into question. In particular, I will try to defend the following three theses. First, semantics is no proper guide to the nature of properties; in order to understand what properties are, we have to look somewhere else. This thesis is supposed to apply to both theoretical semantics as well as to our semantical intuitions based on everyday language use. Second, the phenomenon of self-instantiation is not essential to the nature of properties. From the second thesis we can conclude that finding a solution to Russell's paradox is not the defining task of a theory of properties, in contrast to what most formal workers in the area of property theory seem to assume. And this is the third thesis.

I am pretty sure that many philosophers—especially those working in the field of natural ontology—are willing to accept a position similar to the one that will be presented below. To give an example, Michael Jubien has argued in [11] that *evidence for a good theory of properties must not be confined to the realm of ... semantics*. But also Armstrong, Mellor, and Tooley, to mention only a few, have rejected the view that our understanding of properties mainly rests on semantical considerations. However, their arguments differ essentially from mine. Moreover, I do not share their optimistic view according to which properties are able to exert a strong explanatory power within natural ontology. I would accept the position that it is legitimate to use property talk for reasons of elegance and expressiveness, but I doubt that there are convincing arguments to the effect that they play an *indispensable* role in explanation. But this is a different story which has to be left for another occasion.

To prepare the reader for what is to come, I will begin by laying out the structure of the paper in some detail. In section 2, I recap the argument underlying Russell's paradox and then give a formal presentation of it. In section 3, I sketch some of the escape routes that have been suggested in the literature. The section closes with the claim that any reasonable solution to the paradox should rest on a substantial account of what properties are. This claim raises the problem of finding a proper source for such an account, and it is exactly this problem that initiates the considerations that will be undertaken in the rest of the paper, in the course of which I will present my arguments against the semantic view of properties.

In section 4 it is shown why our linguistic intuitions relying on everyday language use do not solve the problem under discussion. Therefore, we must have recourse to a more theoretical strategy. In order to learn what properties are like, one has to inquire into the explanatory purposes they are supposed to fulfill. This idea, which is quite popular nowadays, will be the topic of section 5. Following common practice, I will identify two theoretical domains over which all applications of properties distribute, namely the domain of theoretical semantics and the domain of natural ontology.

In section 6, I will argue that semantics cannot be taken as the basis for our understanding of the nature of properties. This will turn out to be the most demanding

part of the paper. The argument consists of a number of methodological considerations which finally cut the intimate connection between properties and semantics, and are supposed to prove my first thesis. The second and the third theses follow from the first together with the fact that natural ontology, the other possible source of inspiration, rests on a hierarchical universe of properties. This will be the topic of the final section 7.

2. Russell's Paradox

Russell's paradox has become famous as the most serious challenge every foundational approach to mathematics has to face. Since mathematics lives in an extensional world, Russell's paradox is first of all a problem for set-theorists. However, it also has a twin brother within the realm of properties. Suppose there is a predicate in our language which expresses, for an arbitrary object x, that x does not instantiate itself, and assume further that the property of all things that do not instantiate themselves belongs to the objects we are talking about. We then obtain for this property of non-self-instantiation that it instantiates itself if and only if it doesn't instantiate itself. Therefore, classical logic assumed, we end up in a contradiction.

For the following discussion, it will be useful to have a more formal presentation at hand.[1] So let $\mathcal{L}$ be a first-order language based on an alphabet consisting of an infinite list of individual variables $v_0, v_1, \ldots$; the usual logical symbols $\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \forall, \exists$; an intensional abstractor $[\ldots]$, the (binary) instantiation predicate Δ—where $\Delta t_1 t_2$ expresses that t_1 instantiates t_2—and (optional) a finite number of relation symbols $R_1, \ldots, R_n$.

The terms and formulas of $\mathcal{L}$ are defined by simultaneous recursion, where the only interesting clause concerns the abstraction terms: If φ is a formula and x a variable, then $[\varphi]_x$ is a term which is supposed to denote the *property of being a* φ. To make life easier, I restrict my attention to unary properties.

Russell's paradox is now obtained as follows. Suppose we accept the following two principles—for any $\mathcal{L}$-formula φ with no occurrence of y—where the first fixes the meaning of the abstraction term $[\varphi(x)]_x$, and the second reflects the fact that $[\varphi(x)]_x$ is treated, according to the principles of classical logic, as a *denoting* term:

$$\forall y(\Delta y[\varphi(x)]_x \leftrightarrow \varphi(y/x))^2 \tag{1}$$

$$\exists y(y = [\varphi(x)]_x) \tag{2}$$

Consider the formula $\varphi := \neg\Delta xx$. From (1) we immediately obtain

$$\forall y(\Delta y[\neg\Delta xx]_x \leftrightarrow \neg\Delta yy)$$

[1] Strictly speaking, the formal version will only be used in section 3. So the reader who is exclusively interested in philosophical considerations, may skip the rest of the present section as well as most of section 3.

[2] $\varphi(y/x)$ denotes the formula we obtain from φ by replacing all free occurrences of x with y.

and, by making use of (2), we get

$$\Delta[\neg\Delta xx]_x[\neg\Delta xx]_x \leftrightarrow \neg\Delta[\neg\Delta xx]_x[\neg\Delta xx]_x$$

which forms a contradiction, given the principles of classical propositional logic. Alternatively, one could do without any abstraction terms and use the following instantiation of the general comprehension principle:

$$\exists x\forall y(\Delta yx \leftrightarrow \neg\Delta yy) \tag{3}$$

However, the version with abstraction terms has some advantages and will be taken as a base for the following discussion.

3. Some Formal Solutions

Since Russell discovered his paradox in 1901, a number of escape strategies have been proposed. Feferman [9] supplies a coarse, yet informative, classification by means of the following three slogans:

(A) Restrict the language.

(B) Change the logic.

(C) Restrict the basic principles.

The first slogan reflects the most straightforward route for avoiding the paradox: the idea is just to exclude "dangerous" expressions like $\neg\Delta xx$ or $[\neg\Delta xx]_x$ from the language. In a most consequent way it has been realized in several accounts based on typed languages, the most famous of which is the simple version of Russell's theory of types due to Ramsey. Another important example is Montague's intensional logic IL.

Nowadays, there seems to be wide agreement that for reasons of expressiveness a proper solution to the paradox should be type-free, for a deviating position consult [1] or [10]. Philosophers, linguists, and logicians have argued that many important things one would wish to express are not expressible within a typed language, for details see [4, 6, 7, 16, 23]. To give an example, typed languages are not able to handle *transcendental* predicates, that is, predicates like 'thinks about' or 'loves' which can take arguments of different types, see [5, 14]. Moreover, there are convincing reasons to the effect that a suitable language for property talk should contain variables that range over the whole universe, including properties of all types as well as individuals, see [4, 5, 14]. It is evident that the existence of such uniform variables would violate the usual type restrictions.

But even if we are willing to grant the validity of these arguments—and I feel strongly inclined to do so—we are not justified in drawing the more general conclusion

to the effect that all escape routes relying on syntactic restrictions are damned to fail. For most of the above arguments only address type restrictions, not the general idea of excluding certain terms from the language. However, I agree with many other authors that we should dispense with all kinds of syntactic restrictions.

But again, one has to be careful not to draw the wrong conclusions. That self-referential expressions are admissible in the language, does not imply the *existence* of self-instantiating properties; terms like $[\Delta xx]_x$ need not denote (see below). In general, type-freedom on the linguistic level does not guarantee type-freedom on the ontological level. Having a type-free language by means of which we express facts about entities of a certain kind, is compatible with accepting a hierarchical universe for the same entities.[3]

It is time to turn our attention to the two remaining slogans. Escape strategies that are inspired by (B) or (C) are usually strongly interwoven with each other. Typical examples can be found among property theories which make use of partial logic. Although most of them are driven by (B), they also have to change, as a consequence, (1) and/or (2). Some of these accounts were obtained by adopting techniques which had originally been developed in the area of truth theory, where Kripke's famous paper "Outline of a Theory of Truth" initiated a whole line of research on theories containing their own truth predicate. Property theories like the ones that were proposed in [5, 9, 16, 23] are based on the same model-theoretic intuitions that underly theories of truth developed by Kripke, or Gupta and Herzberger.

Another strategy falling under (B) is based on free logic. The idea is to treat problematic abstraction terms like $[\neg\Delta xx]$ as non-denoting terms. As a consequence, (2) does not apply to them, and, hence, the derivation of Russell's paradox that has been given in section 2 is blocked. As for the underlying logic, the *propositional* part of classical logic can be preserved, only the principles for the quantifiers need to be replaced with their *free* variants. For instance, instead of $\forall x\varphi \rightarrow \varphi(t)$, we take something like $\forall x\varphi \wedge \exists x(x = t) \rightarrow \varphi(t)$. A very natural account following this line results in a property theory which shows great similarity with ZF set theory, where extensionality (and foundation) has been dropped, see [4, 11, 20].

Alternatively, one may wish to treat all abstraction terms as denoting, see [12]. In order to escape the paradox, it is then useful to distinguish two kinds of property-like entities, similar to the set-class distinction known from set theory. The universal quantifier in principle (2) will be restricted to objects of the first kind, whereas terms like $[\neg\Delta xx]$ take their denotations from the second one. A similar account has been proposed by Cocchiarella [8] within a second-order framework. In the same paper the author discusses an alternative theory based on the idea of restricting principle (2) to *stratified* terms.

Although these remarks certainly do not exhaust all strategies that have been proposed in the literature for avoiding Russell's paradox, they should give at least some

[3] In fact, this is the situation we meet in set theory, where the standard version of ZF relies on a first-order type-free language—so that terms of the form $\{x : x = x\}$ or $\{x : x \in x\}$ are well-formed—but the universe of all sets carries a (cumulative) hierarchical structure.

hints as to what is possible. I personally prefer a solution in the style of ZF set theory, but it is not the purpose of this paper to defend any particular solution. Instead, I am aiming at a more basic and conceptual problem. Before we will be able to evaluate and compare different accounts in a reasonable way, we first need a clear understanding of what our standards are supposed to be. The crucial question is: what are the *criteria* by which we should judge a proposed solution to the paradox, and where can we find them?

To begin with, let's consider the reasons that are usually offered in order to show why one should select one solution rather than another. Well, to be somewhat unfair, though I guess it is only "somewhat", it seems to be correct to say that many accounts in the literature are governed by the following rule: preserve the above principles (1) and (2) for as many terms as possible, without ending up in an inconsistency. This is an interesting game, no doubt. In particular, it tells us what is the maximum we could expect. But I don't think that it is a proper game for finding an adequate solution to the paradox.

Instead, I am voting for the following methodological principle, which strikes me as self-evident, and, thus, for which I have no argument to offer, but which will drive the philosophical considerations contained in this paper. For an adequate solution it is not sufficient to avoid the paradox by means of some elegant formal techniques. The solution has to be part of a substantial theory, where the latter is supposed to explicate a clear conception of the nature of properties.

4. Can Everyday Language Use Be our Guide?

At first sight, we seem to have a direct grip on properties. Properties are entities we seem to be well acquainted with. They are part of our everyday world, and, hence, they are deeply involved in our pre-theoretic as well as our scientific reasoning about almost everything. So we should expect ourselves to have a clear comprehension of their nature. At least, we should be able to decide what properties exist, and the latter should bring us very close to a solution to Russell's paradox. At the end of this section, it will turn out that things are not so easy and that our hope was delusive.

Our intimate contact with properties is reflected in our language use. We accept a lot of sentences which deal with properties as true. Of course, we often call them by different names, we say "features", "characters", "traits", and so on, but, nevertheless, what we mean are properties, or special kinds of them. The most characteristic cases of such property talk distribute over two categories of sentences. The first contains sentences by means of which we make universal or existential claims. In more technical terms, these are sentences which *quantify* over properties, like

(a) Peter has the same character traits as his cousin.

(b) There are undiscovered properties tying physical particles to each other.

The second category consists of sentences, where abstract nouns that appear to stand for properties occur in subject position. Famous examples are

(c) Red is a color.

(d) Red resembles orange more than it resembles blue.

Sentences like (a)–(d) were often considered in contexts in which the existence of properties was at stake. In modern times, this problem has often been discussed in recourse to Quine's criterion of ontological commitment. Platonists with respect to properties have utilized sentences like (a)–(d) in order to show that we are committed to the existence of properties, whereas nominalists have invented a number of interesting strategies for avoiding this consequence. Three of the most influential ones are the following:

(i) In the case of quantification, the nominalist gives quantifiers a non-referential, that is in most cases a substitutional interpretation. Examples as (b) put this strategy to a severe test, since they express the existence of properties for which we have no linguistic expression at hand. Still, even though nominalists have offered several escape routes, difficulties remain: In the first place, the nominalist needs a strong argument to show that the sentences under discussion have linguistic expressions as their subjects. For the intuitions to the contrary are rather strong. Second, most substitutional accounts that are able to handle sentences like (b) make use of *possible extensions* of the actual vocabulary. There is some evidence to the effect that a satisfying notion of possible extension of a language depends on the analysis of what a *meaningful* expression is. The difficult task for the nominalist is to provide such an analysis without presupposing the existence of properties, or similar entities.

(ii) The nominalist suggests a suitable paraphrase whose meaning is reasonably similar to the meaning of the original sentence, and whose ontological commitments are acceptable from his own point of view. Here the problems are even more serious than the problems the first strategy had to face: To begin with, since sentence and paraphrase do not have exactly the same meaning, the platonist always seems to be justified to refuse the paraphrase and to insist on a nominalist analysis of the original sentence. Second, there exists no uniform procedure for the nominalist that would enable him to find suitable paraphrases in all cases, and, moreover, there is no argument to the effect that in each single case he can be successful. Third, and even worse, there exists a number of cases where all nominalistic attempts have failed so far; many philosophers are prepared to accept (c) and (d) as cases in point.

(iii) The most promising strategy—which, according to my own view, is the one Quine favored—is driven by the following idea: When we analyze and evaluate ontological commitments, we only have to take into account our best theory(ies), that is, the theory(ies) we use for scientific purposes or other serious work. Hence, the nominalist's task is to show that sentences like (a)-(d) do not occur in her favored theory. It does not matter, whether she supports a position according to which natural language sentences have no ontological commitments at all, and claims that only

sentences in canonical notation do so, or whether she applies the ontological criterion to the former directly. The important point is that sentences which would commit her to the existence of the entities under discussion are not part of the theory(ies) she is willing to accept on the strongest epistemic or ontological standards.

It would be hard work to find out whether one of these strategies could successfully be applied or whether they all must fail. Fortunately, there is no need for doing so. For our problem is not to decide whether properties exist in general, but to find out *which* properties exist, at least, as far as their existence would concern Russell's paradox. So we can even grant that accepting sentences like (a)-(d) as true commits us to the existence of properties. What I am not willing to grant is that from this fact we may draw any interesting conclusions that may help to solve the paradox.

This pessimistic view can be justified as follows. First, we observe that our direct intuitions relying on everyday language use only establish the existence of very 'harmless' properties.[4] They do not enable us to make any decisions with respect to the existence of those properties that are relevant for the paradox. Second, when we try to extrapolate some general existence principles from the obvious cases that are covered by our intuitions, we will end up right back where we started, namely in a paradoxical situation. These two points deserve some comments.

Suppose we are trying to find out whether we (should) accept the existence of a property which instantiates itself. In other words, do our intuitions tell us that the sentence

(e) There is some property that instantiates itself.

is true, as they have told us that (c) and (d) were true? Or is there any argument, resting on our intuitions, to the effect that (e) is true? Perhaps the reader is inclined to give a positive answer by arguing as follows: Take the property of being a property. This is a property which instantiates every property, and, hence, also itself. So there is at least one example of a property which instantiates itself. Is this a convincing argument? Of course not, at best it is an enthymeme, which means that premisses have been suppressed. The most important premiss concerns the existence of the property of being a property.

But why should we accept this premiss? Perhaps it is regarded as self-evident in the sense that the expression "the property of being a property" is well-formed and, hence, has some denotation, which certainly must be a property. But then we are in trouble. To see this, take now the property of being a property that does not instantiate itself. Applying the same principle, this property, call it R, must also exist. But the latter is exactly the property that was responsible for the paradox. In other words, the fact that we trust our intuitions without any restrictions was the reason we got into trouble. So intuitions alone won't give us a way out.

[4]This fits perfectly with the fact that many philosophers who support a sparse conception of properties, and, hence, refuse a simple view on the relation between properties and linguistic items, nevertheless accept the arguments from quantification and abstract reference as (further) evidence for the existence of properties, see, for instance [2]: 58–63.

The reader might reply that our intuitions are all right, and that they tell us that neither the universal property nor R exist, or that the universal property exists, but R does not. The second case looks rather ad hoc to me. We would expect too much from our intuitions, if they should guarantee the existence of the universal property, but not the existence of R. But what is more important here is the fact that the present strategy, just referring to our case-by-case intuitions, gives us no general principles that would help to decide what properties exist.

Taking all this together, I feel justified in drawing the conclusion that everyday language use is no proper guide in our search for an adequate solution to the paradox, and that insights into the nature of properties should be expected from another source.

5. Explanatory Roles

But from what source? Looking for inspiration, it might be reasonable to briefly consider sets. To begin with, what counts there as a satisfying solution to the paradox? At least, the essential condition a proper solution has to fulfill is made out quite easily: It should enable us to use sets as a foundation for the rich field of mathematics. Of course, it would be difficult to formulate this condition in a rigorous manner, but for our purposes it is sufficient to grasp the idea. It is true, there are also some pragmatic restrictions to the effect that the accepted set theory should be elegant and relatively natural—and, perhaps, that at least some of its axioms conform with our intuitions—but these factors only play a subordinate role. And it is also true that we are confronted with deep philosophical questions regarding the application of mathematical theories in sciences. However, as far as I can see, they mainly concern the mathematical theories themselves, and not so much their set-theoretic shape. So what remains is the pure demand that it should be possible to develop mathematics within set theory. And, therefore, to know what sets are is more or less the same as to know what the particular set theory looks like that fits best for this job.

The situation with respect to properties differs in at least three respects. First, and in spite of the negative result from section 4, properties are more deeply involved in our pre-theoretic reasoning about the world than sets are. To use a phrase from Wilfrid Sellars, they are part of our *manifest image* of the world. In fact, this is the general reason why many philosophers are prepared to accept them—beside individuals—as *basic* entities in ontology. Second, there is nothing which can determine our conception of properties in the same way as mathematics determines our conception of sets. Some will object that semantics can do this job, but in section 6 I will show why this is wrong.

Instead, and this is the third and most important point, properties can be used for diverse theoretical purposes; at least, this is the position many philosophers are currently advocating. Properties are supposed to fill various explanatory roles; they help to explain phenomena from very different philosophical areas. In this sense,

arguments for the existence of properties have been constructed as inferences to the best explanation by Swoyer, see [21, 22] and others.[5]

Viewing properties as entities that have been postulated for theoretical purposes, suggests a position according to which their nature is at least to some extent—but only to some[6] —determined by the roles they are supposed to fill. Hence, a proper way to find out what properties are is by investigating these roles.

Several philosophers have made the observation that most applications of properties can be attached to one of the following two domains, natural ontology and (theoretical) semantics. From the perspective of natural ontology, properties serve as the basis of objective similarity in the world, and fix the world's causal structure. Objects resemble or don't resemble each other because they instantiate certain properties, and they have various powers and dispositions due to the fact that they instantiate these properties. Relying on these intuitions, philosophers have assigned properties a central position in their philosophical accounts of scientific realism, causation, dispositions, natural laws and measurement. Within semantics, properties are regarded as ingredients of a theory of meaning. They are used first and foremost as meaning-objects correlated with general terms.

Many philosophers take it for granted that both types of roles have to be performed, the semantical as well as the ontological one, but at the same time they doubt that a single kind of entity is able to perform them all. As a consequence, these philosophers have postulated two kinds of property-like entities. Prominent examples are Bealer's distinction between concepts and qualities, Lewis's distinction between properties and universals, and Putnam's distinction between predicates and properties.

However, whether we should postulate two kinds of properties or not, need not concern us here. It would be sufficient to have at least one clear notion of what properties are like. What will concern us for the rest of the paper are the following two questions:

1. Does semantics give us such a clear notion?
2. What insights do we obtain from the two domains that could enable us to find a reasonable solution to Russell's paradox?

The reader might miss a third question on the list which amounts to whether natural ontology can be a proper guide to the nature of properties. I am sorry that I have to disappoint here, but answering this question is not one of my aims in this paper. So let's turn our attention to semantics!

[5] Of course, the notion of explanation here is highly ambitious. It seems clear that it is different from the notion of explanation that is at work in science, since the latter means first of all causal explanation. Explanations within ontology have more to do with unification or even conceptual analysis. However, it would go beyond the purposes of this paper to discuss this point. For the following, I make the harmless assumption that there is at least one sense of explanation according to which properties are able to undertake explanatory work.

[6] This qualification is essential here, since I regard it as inadequate to *identify* the nature of properties with their theoretical roles. We should not completely ignore the fact that properties are part of our everyday conception of the world. It is exactly this fact which distinguishes them from other abstract objects.

6. Can Semantics Be Our Guide?

The primary motivation for many (formal) theories of properties stems from the desire to develop a new foundation for natural language semantics. There are various semantical phenomena which seem to require the postulation of properties. In particular, properties are wanted for the analysis of predicative expressions.

> The idea that semantics provides a fundamental testing ground for a property theory is based on the assumption that properties are the semantic counterparts of natural language expressions. So, for example, the sentence *John runs* says of John that he has or instantiates the property of running. ([7]: 261)

Other phenomena that appear to imply the existence of properties are self-instantiation, intensionality—properties are postulated in order to get straight the substitution conditions in intensional contexts—and quantification.

Although there is some justification for the above view, from a philosophical point of view it is not sufficient just to postulate the existence of a certain kind of entities, to attach the label 'property' to them, and to correlate them with linguistic expressions. One has to give a full-blown account of what these entities are like, and, especially, how and in what sense they act as *semantic* objects. After all, it is possible to associate many sorts of things with linguistic expressions. The essential question is what makes this association a *semantical* one.

To get a first impression of the whole problem, let's have a short look at the semantics for modal languages. There is wide agreement that possible world semantics has some advantages over its algebraic competitor. It goes without saying that the latter forms an important mathematical tool for investigating modal languages and logics, but it must be questioned whether it forms an explanatory theory of meaning that helps to understand modal talk more properly. For instance, it is not clear what homomorphisms between syntactic and semantic algebras distinguishes from other sorts of homomorphisms, and what qualifies semantic algebras as objects that give meaning to linguistic expressions.

The situation for possible world semantics is rather different. First of all, its ingredients, that is to say possible worlds and individuals, form a suitable basis for intuitive interpretations: many appealing philosophical explanations have been couched in terms of possible worlds. As a consequence, notions like "truth in a world" obtain the status of something that is meaningful to us. That we are justified in viewing the model-theoretic consequence relation as an explication of our informal notion of correct reasoning presupposes this kind of meaningfulness. (That this has to be taken with some care, was shown by Etchemendy.) However, philosophers like Lewis, Plantinga, and others, were still not satisfied. What they tried to do is to develop a full-fledged theory of possible worlds. Though I have some reservations with respect to the whole discussion, I agree with the underlying methodological principle: possible worlds are

only able to satisfy explanatory purposes if we have a clear understanding of what they are.

Transferred to properties, their considerations show that it is not sufficient to use certain structures to interpret some given language and call some elements from these structures properties. If properties should do explanatory work, one has to add a theory which provides at least some information about what their nature is supposed to be. And to characterize their nature is not exhausted by describing the formal features they must have in order to perform their roles as meaning objects. For instance, we need an analysis of the widely accepted fact that the realm of properties is equipped with a kind of algebraic structure similar to the one that underlies the collection of predicates. Otherwise Quine was right, and properties are nothing but mysterious shadows of linguistic expressions.

There certainly exists a very natural reply to the above considerations, which is in full accordance with common practice as it has been described in section 5. Properties are postulated in order to explain and analyze certain phenomena. So they are introduced on purely theoretical grounds, similar to theoretical entities in the natural sciences, like atoms, fields, and quarks. It is, therefore, beside the point to require some characterization of these entities independent of the theoretical roles they are supposed to play. In a first reply to this argument I would like to point out important asymmetries between properties, on the one hand side, and theoretical entities within physics or chemistry on the other side.

First, although theoretical entities in science are postulated on theoretical grounds, there are *empirical* phenomena which give reasons for postulating them, and we usually have empirical procedures and experiments which connect these entities with empirical situations. It must be admitted that these correlations can be rather indirect and very complicated. Nevertheless, they help us to confirm the existence of theoretical entities by empirical evidence, and hence, in a wider sense, the existence of theoretical entities can be regarded as an empirical affair. Properties, in contrast, are established by means of a priori reasoning, at least, as long as their existence is grounded on purely semantical considerations.

Second, it is reasonable to consider the distinction between theoretical and observable entities in science as something flexible that can change in the course of time. As an effect of scientific and technological developments, entities that have been postulated on purely theoretical grounds are becoming observable.

Third, although theoretical entities may differ from observable objects in important respects, with respect to their basic ontological features they are similar. Theoretical entities belong to the same ontological category as observable objects; they are concrete, and, hence, part of the space-time system of our world. In contrast, properties are abstract objects and hence—according to the common view in semantics—they are not situated in space and time.

For the semanticist there are, in principle, two different ways of replying to the above arguments. Either she can contest the disanalogies or she can doubt that they are of any relevance to the point under discussion. Let's begin with the first alternative. It

is probably quite hard to imagine a reasonable position which denies the disanalogies completely, but philosophers like David Armstrong seem to come very close to it. At least they have tried to make them less sharp by arguing that properties—or universals in Armstrong's case—are part of our natural world, and that their existence can only be established by empirical means:

> What properties and relations there are in the world is to be decided by total science, that is, the sum total of all enquiries into the nature of things. ([3]: 8)

I must confess that I have some reservations with respect to Armstrong's position, but they need not concern us here. The only important thing is that his position does not support the semanticist. On the contrary! Since semantics is to a great extent an a priori affair, the naturalness of properties shows, according to Armstrong, that properties are first and foremost entities which are postulated for reasons that are independent of semantical considerations.

More promising for the semanticist seems to be the second alternative. Grant the disanalogies, but doubt their relevance. What is important with respect to properties, and what justifies the analogy with theoretical entities, is the pure fact that properties help to explain a number of things. There is no a priori reason to the effect that such explainers must belong to the category of concrete objects.

Taking this point for granted, the task for the semanticist is then to convince us that it is properties we are really in need of, and not some other kind of abstract entities. To this invitation the semanticist has again two different replies. She can either accept it and supply some reasons in favor of properties, or she refuses it as beside the point, since properties have been *defined* as those entities which fulfill certain roles.

Let's consider the first alternative first. Why should we prefer properties? The most natural answer refers to their individuation principle. Properties are more fine-grained than sets, and even more fine-grained than modalized sets, for instance Lewis's sets of *possibilia*. Two properties can be necessarily co-extensional without being identical. This enables us to make more distinctions within the realm of meaning objects, and it is just these distinctions we seem to be in need of. The best example stems from the analysis of intensionality: the task is to find meaning objects that supply the correct substitution conditions for intensional sentences.

For the sake of illustration, it may be useful to describe the structure of the problem in some schematic way. Given some collection Φ of sentences, we are searching for an equivalence relation $\sim$ on the set of expressions occurring in Φ such that $\varphi(\theta_2)$ follows from $\varphi(\theta_1)$ under the assumption that $\theta_1 \sim \theta_2$, where $\sim$ is defined by means of the *equality* relation between the meaning objects correlated with the θ_i's. The claim is that for intensional sentences properties are the suitable meaning objects.

We should be very sceptical about this claim. That we are really in need of such fine-grained entities is anything but self-evident. To see this, let's elaborate a bit on this point. According to common practice, we can distinguish two kinds of intensional contexts, *modal* contexts and *intentional* contexts, such as propositional

attitudes. For modal contexts properties are not required; this has been observed by various authors. We can, for instance, work with possible worlds and ordinary sets. What about intentional contexts, say belief contexts? Here many philosophers and linguists have argued that fine-grained objects like properties and propositions are indispensable.

However, the situation is not as clear. In many cases the judgement that properties are indispensable rests on an overestimation of the reach of semantics. What we can expect from semantics is that it describes or explicates our conception of *correct* reasoning. It may tell us that if somebody believes φ then, under such-and-such conditions, he should also believe ψ, or, if he is committed to φ, he is also committed to ψ. What semantics cannot do is to describe what people *actually* believe or think. But it is the latter that is often supposed when one argues from fine-grainedness.

On the other hand "often" does not mean "always". So what about those cases that do not rest on the above confusion? Well, I am quite sure that there are no such cases. There is strong evidence to the effect that it is always possible to replace properties by other entities. However, a detailed proof for this claim would go beyond the scope of this paper. Instead, I will try to bring the point home by two further arguments:

First, even if it were desirable to have a proper notion of fine-grainedness, it is fair to say that at present there is none in sight. What we observe is the following dilemma: either the suggested solution doesn't succeed or it pushes properties too close to linguistic expressions so that, again, they are nothing but worthless duplicates of the latter. For an instructive example, consider Bealer's second conception of properties.

Second, suppose there is a chance to improve our situation by finding a suitable notion of fine-grainedness—although I am convinced that this won't happen—why should this prove anything to the benefit of properties? In the end, we can always take a set-theoretic model that is able to perform the roles properties were nominated for. Since, according to the semanticist's account, all we know about properties are their structural features, there is no way to distinguish them from set-theoretic entities that *carry* the same structure.

True enough, the semanticist counters, adding quickly that there is no reason for making this distinction. Taking the second route from above, she identifies properties with any kind of entities that are equipped with exactly the features required for making them suitable for her purposes. And, in her eyes, this is everything that can be said about their nature.

Here we have reached a point where it is hardly possible to convincingly argue against the semanticist's position without giving a detailed account of the methodology of referential semantics, which is certainly something that cannot be done in a twenty-page article, but would require a whole book, or two. So I have to content with a few very general remarks.

The first repeats a point that has already been made several times before. According to my view it is misleading to treat properties as abstract entities that have *exclusively* been postulated for theoretical purposes. Doing so neglects an important difference

between properties and arbitrary formal entities. It is no accident that properties are called properties, and that they stand in the center of our attention. We should not give up the idea that they are part of our manifest image of the world, at least not, if we take them to be entities that fill explanatory purposes within philosophy.

The second point brings us back to the beginning of this section: In order to understand what properties are, we need a clear conception of their role as meaning objects, and how they manage to perform it. But this entails having answers to the following questions:

- How does it come about that properties are correlated with linguistic expressions?
- What makes this correlation a semantical one?
- How is this related to our ability to understand language?

It is hard to imagine how a successful account of these problems could rest on a purely instrumental view like the one our semanticist seems to accept.

If the reader is still not convinced, I have to play my last ace by referring back to the starting-point of our investigation. From Russell's paradox we learned that a simple correlation between linguistic expressions and properties, which is reflected by the comprehension principle, cannot hold without exceptions. So the question that also inspired this paper amounts to what expressions are correlated with properties. Russell once wrote somewhere that what is true is not something that can be decided by the police, and this certainly applies to the existence of properties as well. So a reasonable solution to the paradox cannot be obtained by pure stipulation, but must be part of a substantial account of properties. This maxim was formulated at the end of section 3.

The semanticist who subscribes to a view according to which properties are just entities which have been postulated for theoretical reasons within semantics, and who reduces their nature to their role as denotations of linguistic items, has no tools at hand that would allow her to realize the above maxim. In the end, she is confronted with various formal solutions, but is not able to evaluate them on the ground of some non-formal criteria.

This remark closes the discussion on the relation between properties and semantics. I hope I have accumulated sufficient evidence for a negative answer to the first question from the end of section 5, and, hence, that I have proved the first main thesis of this paper. Semantics doesn't give us a proper notion of what properties are. The latter also implies a negative answer to the second question from section 5. In order to find suitable conditions for a substantial solution to Russell's paradox, we have to look outside semantics.

7. Can Natural Ontology Be Our Guide?

In section 5 I identified two main domains from which we could expect insights into the nature of properties that may lead our search for a solution to the paradox, semantics and natural ontology. The foregoing section proved the first to fall short of our expectations, so that we have to pin our hopes on the second domain. However, it will turn out that natural ontology is unsuitable for our purposes as well, even if for completely different reasons.

We begin with the observation—and in a sense we already could stop here—that all properties operating in natural ontology find their place in a universe with a *hierarchical* structure. Employing a language of construction, this universe can be described as follows: At the bottom of the hierarchy, we start with a collection of individuals. On the first level we take (all) properties and relations that instantiate those individuals. On the second level we take (all) properties and relations instantiating properties and relations from the first level as well as individuals, and so on.

This may remind one of the cumulative hierarchy of sets, and, indeed, this association is completely justified. But in contrast to the situation in set theory, it would require some further work to develop the above picture into a proper construction. In particular, in the case of properties it is not self-evident what exactly the properties (and relations) are that inhabit a successor level, say level $n+1$. For sets the situation is obvious, we just take all members of the power set of the union of the foregoing levels, but in the case of properties things are more complicated. Certainly, we can require that for every set consisting of individuals and properties from levels lower than $n+1$, there exists a property on level $n+1$ which has this set as its extension. But since we lack extensionality, this property need not be unique. Fortunately, these subtleties can safely be ignored for the following.

What we shouldn't ignore is the fact that for the purposes of natural ontology we only need a small part of this hierarchy, although it is highly contested which part exactly. Even philosophers who refuse to establish the existence of properties on a priori reasons have nevertheless set up some general principles that effectively impose strong restrictions on the structure of the universe of properties. For instance, some have argued in favor of a principle of order invariance which means that all entities instantiated by a particular property must be of the same level. This condition pushes the hierarchy closer to the universe that underlies Russell's simple theory of types. And although Armstrong, for instance, refuses this principle, he still accepts a weak principle of order invariance to the effect that the entities contained in an n-tupel instantiated by a particular relation are all of the same level. Moreover, in [3] he introduces a principle which excludes properties and relations of a level higher than 2, which yields a very flat hierarchy.

But no matter what principles are accepted, there seems to be common agreement that everything we need in natural ontology lives in the hierarchical universe. As a consequence, the phenomenon of self-instantiation plays no role at all. Philosophical

accounts of scientific realism, causation, laws, disposition and measurement are not in need of self-instantiating properties. Bad news for the latter!

But many current philosophers are inclined to go even further. According to their view, it is not only the case that we can do without self-instantiating properties, but also that our naturalistic conception of properties *excludes* the possibility of their existence. Properties are regarded as the entities that make the differences in the world, and this world is nothing more than the *natural* space-time system. Moreover, to make differences in the world means to contribute to the causal structure of the world.

> The essential nature of a property consists in its potential for constituting to the causal powers of the things that have it, and that properties are identical just in case they make, in all possible circumstances, the same contribution to causal powers. ([19]: 291)

Since it is incomprehensible that self-instantiating properties, which are just formal entities, could have any causal effects, self-instantiating properties cannot exist at all. So the argument goes.

But even if we do not accept this strong naturalistic position, the phenomenon of self-instantiation has to give up its dominant role in our thinking about properties, which was the second main thesis that should be established in this paper. This conclusion has far-reaching consequences for the status of Russell's paradox, and for formal accounts of properties in general. In particular, if philosophers like Armstrong, Shoemaker, and Tooley are right, and our knowledge about properties is an empirical affair, the whole project of developing formal theories of properties that are driven by the idea of avoiding the paradox turns out to be on the wrong track.

It may still be interesting to do formal work for properties, but the character of this research will have to change. What is now wanted are flexible formal frameworks which enable us to describe the fine-structure of the universe of properties as it is investigated by philosophers inside and outside natural ontology. This new orientation may further have the pleasing side effect, that it initiates more collaboration between two groups of researchers—logicians and philosophers—working on problems related to properties. At present, a fruitful strong cooperation is still missing.

Acknowledgement. I would like to thank Johannes Haag, Volker Halbach, David Hyder, and an anonymous referee for helpful comments on an earlier version of this paper.

References

[1] Anderson, C. Anthony: 1987. Bealer's quality and concept. *Journal of Philosophical Logic* 16: 115–64.

[2] Armstrong, David M.: 1978. *Universals and Scientific Realism, vol. I. Nominalism and Realism*. Cambridge: Cambridge University Press.

[3] Armstrong, David M.: 1978. *Universals and Scientific Realism, vol. II. A Theory of Universals*. Cambridge: Cambridge University Press.

[4] Bealer, George: 1983. *Quality and Concept*. Oxford: Oxford University Press.

[5] Bealer, George and Uwe Mönnich: 1989. Property theories. In: D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic IV*, Dordrecht: Kluwer, 131–251.

[6] Castañeda, Hector-Neri: 1977. Ontology and grammar: I. Russell's paradox and the general theory of properties in natural language. *Theoria* 42: 44–92.

[7] Chierchia, Gennaro and Raymond Turner: 1998. Semantics and property theory. *Linguistics and Philosophy* 11: 261–302.

[8] Cocchiarella, Nino B.: 1992. Cantor's power-set theorem versus Frege's double-correlation thesis. *History and Philosophy of Logic* 13: 179–201.

[9] Feferman, Solomon: 1984. Toward useful type-free theories, I. *Journal of Symbolic Logic* 49: 75–111.

[10] Fine, Kit: 1977. Properties, propositions and sets. *Journal of Philosophical Logic* 6: 135–91.

[11] Jubien, Michael: 1989. On properties and property theory. In: G. Chierchia *et al*. (eds.), *Property Theories, and Semantics, vol. I*, Dordrecht: Kluwer, 159-75.

[12] Lemmon, Edward J.: 1963. A theory of attributes based on modal logic. *Acta Philosophica Fennica*: 95–122.

[13] Lewis, David: 1983. New work for a theory of universals. *American Philosophical Quarterly* 61: 343–77.

[14] Menzel, Christopher: 1993. The proper treatment of predication in fine-grained intensional logic. *Philosophical Perspectives* 7: 61–87.

[15] Orilia, Francesco: 1991. Type-free property theory, exemplification and Russell's paradox. *Notre Dame Journal of Formal Logic* 32: 432–47.

[16] Orilia, Francesco: 2000. Property theory and the revision theory of definitions. *Journal of Symbolic Logic* 65: 212–246.

[17] Pollard, Stephen and Norman M. Martin: 1986. Mathematics for property theorists. *Philosophical Studies* 49: 177–86.

[18] Reinhardt, William N.: 1980. Satisfaction definitions and axioms of infinity in a theory of properties with necessity operator. In: A. I. Arruda *et al*. (eds.), *Mathematical Logic in Latin America*, Amsterdam: North-Holland, 267–303.

[19] Shoemaker, Sidney: 1980. Properties and inductive projectability. In: L. J. Cohen and M. Hesse (eds.), *Applications of Inductive Logic*. Oxford: Clarendon Press, 291–320 .

[20] Sturm, Holger: *Properties in Natural Ontology: An Axiomatic Approach*, forthcoming.

[21] Swoyer, Chris: 1999. How ontology might be possible: explanation and inference in metaphysics. *Midwest Studies in Philosophy* 23: 100–131.

[22] Swoyer, Chris: 2000. Properties. *Stanford Encyclopedia of Philosophy*, 42ff.

[23] Turner, Raymond: 1987. A theory of properties. *Journal of Symbolic Logic* 52: 455–72.

Fachgruppe Philosophie
Universität Konstanz
78457 Konstanz
Germany

E-mail: Holger.Sturm@uni-konstanz.de

The Many Lives of Ebenezer Wilkes Smith

Vann McGee

Abstract. Peter Unger's "Problem of the Many" is anticipated in Bertrand Russell's 1923 article, "Vagueness." Unger's argument depends on principles about the composition of bodies that, while highly plausible, are not beyond dispute. Russell's variant does not depend on the same principles, looking instead to the lack of a precise answer to the question when a human life begins or ends. By foreclosing a possible escape from Unger's conclusion, Russell's version of the argument provides, it is argued, persuasive evidence for the inscrutability of reference.

Science tells us that a material body, such as the mountain Kilimanjaro, is made up out of atoms. Precisely which atoms will vary from time to time, but at each moment since the mountain's creation there have been atoms that made up Kilimanjaro. (There have been various subatomic odds and ends as well, but we may ignore them. Also, let me take it for granted that, if Kilimanjaro contains part of an atom, it contains the whole thing.)

That there is a set consisting of the atoms that make up Kilimanjaro is perhaps not obvious. We are mainly used to talking about sets in the context of pure mathematics, where the sets we discuss are, by and large, the extensions of fully precise predicates. "Is an atom that is now part of Kilimanjaro" is an imprecise predicate, and we might be inclined to suppose that such predicates lack extensions. However, science teaches us that there is now at least one atom in Kilimanjaro and that there are fewer than 10^{100} of them, and these two facts[1] entail "$(\exists\, set\, x)(\forall y)(y \in x \leftrightarrow y$ is an atom that is now part of Kilimanjaro)," with the aid of principles of set construction so rudimentary that no one who is willing to abide talk about sets at all is going to deny them:

$(\forall x)(\exists\, set\, y)(\forall z)(z \in y \leftrightarrow z = x)$. *Unit sets.*

$(\forall\, set\, x)(\forall\, set\, y)(\exists\, set\, z)(\forall w)(w \in z \leftrightarrow (w \in x \vee w \in y))$. *Pairwise unions.*

Saying which atoms are elements of $\{x \mid x$ is an atom that is now part of Kilimanjaro$\}$ is no easy endeavor. There are quite a few atoms that unmistakably are parts of Kilimanjaro, and quite a few more that unmistakably aren't, but there is no sharp geological border where Kilimanjaro leaves off and the surrounding countryside begins.

[1] We understand "There are at least n" and "There are fewer than n" to be cashed out in terms of quantification theory. "There is least one F" is rendered "$(\exists x)Fx$." "There are at least $n+1$ Fs" is "$(\exists y)(Fy \wedge$ there are at least n Fs other than $y)$." "There are fewer than n Fs" is "$\sim$ There are at least n Fs."

Not every way of partitioning the atoms that aren't either clearly inside or clearly outside Kilimanjaro results in a plausible candidate for $\{x \mid x$ is an atom that is now part of Kilimanjaro$\}$. For example, we would expect the atoms that make up Kilimanjaro to be contiguous. This is so because Kilimanjaro is a material body, and disconnected clumps of atoms don't make up a material body. Taking such considerations into account, however, we still find quite a large number of smooth, simple closed curves that enclose a plausible candidate for $\{x \mid x$ is an atom that is now part of Kilimanjaro$\}$, with no reason to suppose that one of the enclosed sets is a better candidate than another.

Every simple closed curve through the Tanzanian countryside that doesn't cut through any atoms[2] divides Tanzania into two parts, the part inside the curve and the part outside.[3] In particular, each of our candidate borders of Kilimanjaro encloses part of Tanzania. Different curves enclose different parts of Tanzania, at least if they differ enough so that one contains an atom that the other excludes. Kilimanjaro is part of Tanzania, but there doesn't appear to be anything, either in the geology of Kilimanjaro or the use of "Kilimanjaro" that determines precisely which part it is. On the contrary, there appear to be countless billions of parts of Tanzania that are all perfectly good candidates for what the name "Kilimanjaro" refers to.

This is a version of Peter Unger's "Problem of the Many," which has emerged as an immensely powerful argument for the inscrutability of reference. It is an argument of breathtaking breadth. It applies not just to names of mountains but also to names of people, countries, battles, bridges, and galaxies. It applies to demonstratives, as seen, for example, when Ernest points toward a snow-capped mountain and says, "That is the tallest mounting in Africa." It applies to definite descriptions, like "the tallest mountain in Africa," and it applies to count nouns as well as singular terms. The geography books tell us that there is exactly one mountain at 3.07° S., 37.35° E., but there isn't anything that picks out exactly one of the individuals at that location as the local referent of "mountain."

Unger's problem has deep consequences. Folk semantics supposes that, in order for a sentence like "The ice pack atop Kilimanjaro is being depleted by global warming" to be true, speakers' usage (broadly understood) has to pick out a unique individual as the referent of "Kilimanjaro." Unger's argument shows that this demand cannot be met. The consequences for our commonsense understanding of *de re* indirect speech reports and propositional attitude attributions are even more disturbing. In order for Ernest to have a *de re* belief regarding Kilimanjaro that it is the tallest mountain in Africa, there need to be some features of his mental state, and of the way his mental state is situated in his body and environment, that single out a particular individual as the thing his belief is about. Or so it would seem; but the problem of the many shows that there are no such features.

The thesis that the problem of the many should be thought of as evidence for the inscrutability of reference, rather than, as Unger himself likes to think of it, as evidence

[2]We exclude candidate borders that cut through atoms because of our assumption that the constituents of Kilimanjaro are whole atoms.

[3]This is the Jordan curve theorem.

that there are no people or mountains, has been developed by Brian McLaughlin and me at some length elsewhere [18]; there is no need to repeat the discussion here. Here I am interested in the origin of the problem. The problem of the many was discovered by Peter Unger,[4] but it seldom happens that an important philosophical idea springs up out of nowhere. Quite often, a thesis developed by a contemporary philosopher is anticipated in the work of an earlier philosopher. With surprising regularity, that philosopher is Russell. So it is with the problem of the many, an early version of which appears in Russell's 1923 paper, "Vagueness" [22].

The example Russell uses is the name of a man, Ebenezer Wilkes Smith, and it relies on the observation that there isn't an exactly determined point at which Smith's life begins or ends to conclude that the name lacks precision. Although Russell expends only a few lines on the puzzle ([22]: 63), the version he presents is interestingly different from Unger's. As we shall see, there are a couple of tiny cracks in Unger's argument, not enough to raise a realistic hope that one could escape from Unger's conclusion, but enough at least to suggest the possibility. Russell's version doesn't have cracks at the same places.

In addition to the problem of the many, other ideas of contemporary currency are anticipated in Russell's "Vagueness," in particular, the idea that the source of the paradoxes of vagueness is the use of classical logic. We read, "All traditional logic habitually assumes that precise symbols are being employed. It is therefore not applicable to this terrestrial life, but only to an imagined celestial existence" ([22]: 65). It's surprising to hear Russell say this, because it's a view one expects to hear from someone who wants to downplay the costs of employing a nonclassical logic for use with vague terms, by treating appearances of vagueness as exceptional special cases. Russell, however, is keenly aware of the ubiquity of vagueness, noting (see [22]: 63) that even our most earnest efforts at scientific rigor succeed only in reducing, not eliminating, the imprecision in our speech. Nearly every term of our language has actual or potential borderline cases, so that, if classical logic requires full precision, we exceed its jurisdiction every time we step outside the realm of pure mathematics. This, in turn, means that rigor forbids us to apply classical arithmetic and classical geometry, which rest atop classical logic, in even the most innocuous counting and measuring situations. We should resist this outcome if we can, for classical applied arithmetic has proven to be among the most reliable as well as the most successful of human scientific endeavors, and we should be reluctant to give it up just to satisfy the demands of semantic theory.

Russell argues for the imprecision of logic on the grounds that the logical operators are vague, since they are defined in terms of "true" and "false," and "true" and "false" are vague. This line of argument is unconvincing. It could happen that a term we use to formulate a definition has borderline cases as well as clear cases, but that the clear cases are sufficient to pin down precisely the referent of the defined term. Indeed, this is the case with the logical connectives, at least according to classical, bivalent

[4] David Lewis [14], Peter van Inwagen [28], and Samuel Wheeler [29] should also be mentioned.

semantics. The fact that some sentences containing "$\vee$" lack determinate truth-values doesn't show that "$\vee$" is vague, just as the fact that "the greatest prime factor of the smallest large integer" lacks a well-defined value doesn't show that the function that takes an integer > 1 to its largest prime factor is imprecisely defined. Indeed, we can show that the largest-prime-factor function is specified precisely (assuming, for argument, that Arabic numerals are precise), by noting that each possible input to the function is rigidly denoted by a precise name, and each result of substituting such a precise name into "the largest prime factor of__" uniquely determines a precisely nameable output. So any imprecision that occurs in a phrase of the form "the largest prime factor of__ " must be the fault of the term that fills in the blank. The situation is similar with the truth function "$\vee$." We can find both sentences that are unequivocally true and sentences that are unequivocally false to plug into the blanks in "__$\vee$__," and the result of doing so invariably yields a sentence that is either unequivocally true or unequivocally false.

The grievance against classical, bivalent semantics is that it doesn't account for vagueness. In insisting that, no matter what amount of hair Harry has on his head, "Harry is bald" is either true or false—a conclusion that follows logically from "'Harry is bald' is true if and only if Harry is bald" and "'Harry is bald' is false if and only if Harry is not bald"—classical semantics doesn't deny that Harry is a borderline case of "bald." Bivalence only implies precision if we assume some principle that ensures that, if "Harry if bald" is true, then the thoughts and practices of the community of speakers, together with the configuration of hairs on Harry's head, make it true, and no such principle is to be found in classical semantics as presented in Tarski's *Wahrheitsbegriff*. The trouble with classical semantics isn't that it denies vagueness but that it ignores it, treating vague and precise predicates just the same way.

To remedy this deficiency while maintaining classical logic, three approaches have been proposed:

(1) To maintain classical semantics, just as it is, but to add a new operator "*Determinately*," so that, if Harry is a borderline case of "bald," Harry will be either bald or not bald, but he won't be either determinately bald or determinately not bald, and "Harry is bald" will be either true or false, but it won't be either determinately true or determinately false. The classical rules of inference are both truth-preserving and determinate-truth-preserving.

(2) To allow truth value gaps, so that, if Harry is a borderline case of "bald," "Harry is bald" will be neither true nor false. The classical rules are truth-preserving.

(3) To employ a many-valued logic, with truth values taken from a complete[5] Boolean algebra, and logical operators understood compositionally, so that,

[5]Completeness is a stronger hypothesis than we really need. We assume it because it makes the action of the quantifiers so easy to describe. Let's call a function taking individuals to truth values a unary *Fregean concept*. The existential quantifier designates the operator taking a unary Fregean concept to the supremum of its range, while the universal quantifier takes it to the infimum of its range. Completeness ensures that these suprema and infima are defined. The quantifiers take $n + 1$-ary Fregean concepts to n-ary Fregean

> for example, the truth value of a disjunction is the sum of the truth values of its disjuncts. If Harry is a borderline case, "Harry is bald" will have a truth value intermediate between 1 (truth) and 0 (falsity). If Harry and Jim are both borderline cases of "bald," but Harry has markedly less hair than Jim, then the truth value of "Harry is bald" will be greater than that of "Jim is bald." The truth value of the conclusion of a classically valid argument will be greater than or equal to the infimum of the truth values of its premises.

The key idea that we need to implement any of these proposals is that the usage of a community of speakers doesn't single out a unique intended model of the language. Instead, it picks out a family of acceptable models. "Models" here are ordinary classical models, as described, for example, in [2]. A model assigns a set of n-tuples to each n-place predicate. For arbitrary models, this set can be assigned arbitrarily, but for acceptable models, the choices are restricted. People who are definitely bald will be in the set assigned to "bald" in each acceptable model. People and things that are definitely not bald will be outside the set an acceptable model assigns to "bald." The range of intermediate cases—what Russell calls the "penumbra"—will be adjudicated by different acceptable models in different ways, subject to constraints like the following: Any acceptable model that puts Jim into the set assigned to "bald" and puts ⟨Harry, Jim⟩ into the set assigned to "has less hair (on his scalp) than" will put Harry into the set assigned to "bald."

On the bivalent approach, a sentence is determinately true if and only if it is true in all the acceptable models. On the truth-value-gaps approach, such sentences are simply true. The many-valued approach takes the truth value of a sentence to be the set of acceptable models in which it is true, defining the sum of two such sets to be their union, the product to be their intersection, and so on.

Each approach permits us to introduce an operator "*Determinately*," understood so that *Determinately* ϕ says that ϕ is true in every acceptable model.[6] More generally, an open sentence with n free variables has an *extension*, consisting of n-tuples that definitely satisfy the formula; an *antiextension*, consisting of n-tuples that definitely fail to satisfy the formula; and a penumbra. n-tuples in the extension of ϕ satisfy *Determinately* ϕ. "*Determinately*" acts a lot like "□," with acceptable models playing the role of possible worlds. In particular, *precise names*—names that denote the same thing in each acceptable model—play a role analogous to that played by rigid designators in modal logic. Precise names can be substituted into the scope of "*Determinately*," just as proper names can be substituted into the scope of modal operators. If c is a precise name of a, then *Determinately* $\phi(c)$ holds just in case a is in the extension of ϕ. Precise names can be used to instantiate universal generalizations containing "*Determinately*"; imprecise names cannot. Thus, using "It is determinate whether ψ"

concepts in an analogous way. The identity sign designates the binary Fregean concept taking unlike ordered pairs to 0 and pairs of the form $\langle a, a\rangle$ to 1.

[6] Because the distinction between acceptable and unacceptable models is imprecise, there will be cases in which it is indeterminate whether it is indeterminate whether ϕ. As we ascend the Tarski hierarchy of languages, we may expect to encounter vagueness all the way up; see [17]: appendix.

as an abbreviation for (*Determinately* $\psi \vee$ *Determinately* $\sim \psi$),

> $(\forall x)(\forall y)(x$ and y are integers $> 1 \rightarrow$ it is determinate whether y is the largest prime factor of $x)$

and

> The smallest large integer is an integer > 1

do not entail:

> $(\forall y)$(it is determinate whether y is the largest prime factor of the smallest large integer).

Everyday names of ordinary objects—names like "Kilimanjaro," "the tallest mountain in Africa," and "Ebenezer Wilkes Smith"—are imprecise. This was shown by Gareth Evans in a famous one-page article [3].[7] The argument is highly compressed and requires quite a bit of reconstruction,[8] but I'll try my best. Let a_1, a_2, a_3,..., a_N be a list by precise names of all the Kilimanjaro candidates, that is, of all the mountainlike individuals that we get by adjudicating the disputed atoms on Kilimanjaro's border in every possible way. Thus we have "Kilimanjaro $=$ $a_1 \vee$ Kilimanjaro $= a_2 \vee$ Kilimanjaro $= a_3 \vee \cdots \vee$ Kilimanjaro $= a_N$," because the a_is name all the candidates. We also have the law of logic "$(\forall x)(\forall y)(x = y \rightarrow$ *Determinately* $x = y)$", derived from "$(\forall x)$ *Determinately* $x = x$" and the identity axiom "$(\forall x)(\forall y)(x = y \rightarrow$ *Determinately* $x = x \leftrightarrow$ *Determinately* $x = y)$." By hypothesis, each of the a_is is precise; so we can substitute a_i into the scope of "*Determinately*" to obtain "$(\forall x)(x = a_i \rightarrow$ *Determinately* $x = a_i)$." Assuming, for *reductio ad absurdum*, that "Kilimanjaro" is precise, we can substitute again to derive "(Kilimanjaro $= a_i \rightarrow$ *Determinately* Kilimanjaro $= a_i$)." Thus we obtain the absurd conclusion,"*Determinately* Kilimanjaro $= a_1 \vee$ *Determinately* Kilimanjaro $= a_2 \vee$ *Determinately* Kilimanjaro $= a_3 \vee \cdots \vee$ *Determinately* Kilimanjaro $= a_N$."

"a_i= Kilimanjaro" is indeterminate. "=" is precise—it denotes the same set of pairs in every acceptable model (let us agree to ignore indeterminacy at the edge of the domain of quantification). "a_i" is precise, by hypothesis. So the only place the indeterminacy can come from is "Kilimanjaro."

The argument assumes that there are finitely many Kilimanjaro candidates and that each of them can be named precisely, but these assumptions are inessential. If the collection $\mathcal{C}$ includes all the Kilimanjaro candidates, we have "$(\exists x)(x \in \mathcal{C} \wedge$

[7] In regarding Evan's argument this way, I am taking liberties with the text, inasmuch as Evans himself takes his topic question to be "Can there be vague objects?" The reason I prefer to interpret the paper so that its central issue is the vagueness of names, rather than the vagueness of objects is that I am inclined to agree with Russell in regarding the thesis that there are vague objects, if it is taken literally, as a suggestion we can dismiss out of hand. "Vagueness and precision alike," he says ([22]: 62), "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 there are, and there is an end of it." See [8]: ch. 1.

[8] In this, I was helped by [13], [1], and [6]. See also [16].

Kilimanjaro $= x$)". Assuming that "Kilimanjaro" is precise, we can substitute it into the scope of "*Determinately*" in a law of logic to obtain "$(\forall y)$(Kilimanjaro $= y \rightarrow$ *Determinately* Kilimanjaro $= y$)," from which we derive the absurd conclusion "$(\exists x)(x \in \mathcal{C} \wedge$ *Determinately* Kilimanjaro $= x$)."

The reconstructed Evans argument shows that, if there are many candidates for what "Kilimanjaro" refers to, then "Kilimanjaro" is not precise. The only way "Kilimanjaro" can be precise will be if there is but a single candidate, so that, of the various things[9] that are present at 3.07° S., 37.35° E., there is only one that closely resembles a mountain. We have already seen that there are numerous apt candidates for $\{x|x$ is an atom that is now part of Kilimanjaro$\}$. The *single-candidate thesis* would have it that exactly one of these many sets constitutes an individual;[10] or if there is a sense of "constitute" in which many of the candidate sets constitute individuals, it is sharply different from the way sets of atoms constitute mountains. It is determined that exactly one set constitutes an individual (in the right sort of way), but it isn't determined which set that is; this is because the constitution relation is vague. If B_1 and B_2 are two different candidates for Kilimanjaro's border, then there is a sense of "constitute" in which the stuff inside B_1 constitutes an individual, and a different sense in which that same individual is constituted by the stuff inside B_2, but there is no sense of "constitute" (or, at least, no sense of "constitute" that closely resembles the sense in which a mountain is constituted by the stuff inside it) in which both the stuff inside B_1 and the stuff inside B_2 constitute individuals.[11] If our understanding of the relation of parts and wholes is governed by the principles of mereology, as developed by Leśniewski [10] and by Goodman and Leonard [5], and if we suppose that a mountain is composed of a lot of tiny parts,[12] then the single-candidate thesis

[9] I intend to use the words "thing," "entity," and "individual" as widely as possible, so that everything is thing, everything is an entity, and everything is an individual, and there isn't anything that isn't a thing, an individual, and an entity.

[10] I suppose one ought not really to say that a set of atoms constitutes a mountain, but rather that the members of the set constitute a mountain; but the extra verbiage is tiresome.

[11] When I speak of the "single-candidate view," the position I have in mind is one that uses the hypothesis that there aren't any other personlike things in the immediate vicinty of Ebenezer Wilkes Smith to evade the nasty apparent consequences of Unger's problem. By "consequences" here I mean consequences by ordinary classical logic. To respond to the paradoxes of vagueness by abandoning classical logic seems like a bad idea, for reasons given in ([30]: ch. 4), and ([15]: ch. 4); and to respond by both adopting single-candidate metaphysics and abandoning classical logic seems like overkill. I may reasonably be accused of battling an imaginary opponent here, inasmuch as no one I know of has explicitly espoused the single-candidate view, although, in different ways, [31], [7], and [8] have come close. Conversations with Jónsson have been especially useful in helping me understand the view's ramifications.

[12] These tiny parts won't be atoms, in the chemist's sense of "atom." (They might or might not be atoms in the mereologist's sense.) If A is a chemist's atom that is now part of Kilimanjaro (according to our everyday way of speaking about parts) but that, due to erosion, will not be part of Kilimanjaro a million years from now, then a million years from now, the mereological sum Kilimanjaro$+A$ will include a part that won't be part of Kilimanjaro, so Kilimanjaro$+A$ is distinct from Kilimanjaro. This was first pointed out to me by Kit Fine, although I don't know whether it was he who first noticed it. The assumption that Kilimanjaro is composed of tiny parts is nontrivial. Like everyone else, Judith Thomson believes that Kilimanjaro is made up out of tiny parts, if "part" is understood in the everyday sense, but she thinks that, in the mereologist's technical usage, Kilimanjaro doesn't have any proper parts. She accepts the axioms of mereology, although

will founder at the outset. Kilimanjaro is the mereological sum of its tiny parts, and the axioms of mereology ensure that there are plentiful other mereological sums of tiny parts that differ from Kilimanjaro only very slightly. But when mereologists talk about parts and wholes, they employ a technical usage that diverges from ordinary English, most obviously in the fact that ordinary usage allows us to speak of a thing losing old parts and acquiring new ones, so that an atom might be part of Kilimanjaro now but, due to erosion, not a part of it a million years from now. Mereology makes no such allowance.

Mereology is sufficient, but it isn't necessary to get the problem of the many going. Unger's argument will go through on almost any conception of wholes and parts, as long as it provides that familiar objects like Kilimanjaro and Ebenezer Wilkes Smith are made up of many tiny parts and that we get wholes by putting together tiny parts in fairly arbitrary ways. We can thwart Unger's argument by adopting the single-candidate hypothesis, according to which the ways in which tiny pieces can go together to make wholes is highly restricted.

Folk semantics has it that, whenever the sentence obtained by substituting the singular term τ into "x exists" is true, τ refers, and, moreover, in order for τ to refer, the activities of the community of speakers have to single out a particular object as τ's referent. No one who acknowledges semantic indeterminacy[13] (and accepts classical logic) is going to uphold these principles without restriction. "The shortest tall man" (assuming no ties in height), "the smallest large number," and "$\{x|x$ is an atom that is currently part of Kilimanjaro$\}$" are counterexamples. The single-candidate thesis can save proper names from indeterminacy, perhaps, but, inasmuch as we are going to have indeterminately referring singular terms anyway, if that was all the thesis got us, it would seem a lot of effort for little gain.

The real attraction of the single-candidate thesis is that it promises to enable us to uphold the natural way of drawing the distinction between *de dicto* and *de re* propositional attitude attributions and indirect speech reports. On a *de dicto* or *notional* reading of "Ralph believes someone is a spy," the content of Ralph's belief is an existential claim that can be symbolized "$(\exists x)x$ is a spy." On a *de re* or *relational* reading, some particular individual is picked out as the object of Ralph's belief; $(\exists x)$(Ralph believes that x is a spy). See [20]. One popular account has it that the content of Ralph's belief is a *singular proposition*, obtained by supplying a suspected espionage agent as argument to the *propositional function* described by the open sentence "that x is a spy."

Appealing as this account may be, it cannot be correct, if Unger is right about inscrutability of reference, for there isn't anything in Ralph's belief state (even when we take account of how Ralph's beliefs are situated in his body and how his body is

she doesn't think the theory is any use in describing the parts of bodies; bodies are mereological atoms, in her view. Mereology proves its worth in describing the parts of events; see [24].

[13]Timothy Williamson [30] denies that there is semantic indeterminacy. He thinks linguistic usage establishes a sharp boundary for every meaningful term, although speakers are often unable to discern the boundary.

situated in his environment) that picks out a unique individual as the individual his belief is about. Suppose that Ralph is correct in the *de dicto* belief that there is, within the relevant context, one and only one man in a brown hat, and that that man is a spy, and suppose further that the germane brown-hatted man is Bernard J. Ortcutt. Even so, there isn't anything in Ralph's belief state that ensures that the content of his belief is the true singular proposition we get by supplying Ortcutt as an argument to "that x is a spy," rather than one of the innumerable false propositions we get by supplying rival Ortcutt-candidates. (Keep in mind that, on the standard view, it is Ortcutt himself, and not Ortcutt under a description or Ortcutt under a mode of presentation that is a component of the true singular proposition.) Ralph points an accusatory finger and says, "He is a spy," but there isn't anything in Ralph's gesture or its concomitant mental state that would make it correct to say (if *de re* indirect speech reports require determinate referents) that Ralph said of Ortcutt that he was a spy.

As a device for salvaging the natural account of *de re* propositional attitude attributions and indirect speech reports, the single-candidate hypothesis has the feel of wishful thinking. The problem of the many is perplexing to our commonplace ideas about how language is linked to the world, and we can avoid its discomforts if we assume that nature is cooperative enough to provide exactly one object for each term that we ordinarily think of as a denoting proper name. But why should we think nature would be so kind?

Lewis ([12]: 211–123) makes a complaint against the single-candidate doctrine, although, regrettably, he doesn't fill in the details. If we start with a mountain and repeatedly perform the operation of removing a single atom from its surface, we'll eventually obtain something that is clearly not a mountain. This implies by logic that, at some stage, we have removed a single atom from a mountain and obtained something that's not a mountain. It does not, however, imply that "mountain" has a sharp border. At a certain stage, we remove an atom from a mountain, and we get something that, while not a mountain, is so very much like a mountain as to be indistinguishable from one. According to the single-candidate view, if we continue removing atoms one at a time from what used to be our mountain, eventually we remove a single atom from a thing, and we're left with a bunch of atoms that don't constitute anything at all. This time, however, we are stuck with a sharp boundary. The alternative—that removing the atom leaves us with something that, while not existing, enjoys a status so close to being as to be nearly indistinguishable from being—is unintelligible.

Let us return to Ebenezer Wilkes Smith. Russell reflects that neither birth nor death (even what we call "sudden death") is an instantaneous process, so that there will be stages at the beginning and the end of Smith's life at which it isn't determined whether or not the name "Ebenezer Wilkes Smith" is applicable. So far, we can explain what's going on by pointing out that the word "applicable" is vague. But now, imagine a situation in which I point at the figure in Smith's bed, intending to point to Smith, and I say, "That is Ebenezer Wilkes Smith." If, unknown to me, Smith is *in extremis*, it may be that I speak at a stage at which it is indeterminate whether what I say is true. But now consider. On the single-candidate model, demonstratives, when used with the

level of precision ordinarily thought to constitute a successful act of demonstration, aren't vague. "Ebenezer Wilkes Smith" isn't vague. Identity, we have seen already, isn't vague. The sentence is indeterminate, yet there isn't any place in the sentence where indeterminacy creeps in.

If we think of a person as made up of temporal stages, we can assimilate the "Ebenezer Wilkes Smith" example to the "Kilimanjaro" example. It is indeterminate whether the present moment's temporal slice of the object I'm pointing to on the bed is part of Smith. But the example doesn't depend on the temporal-stage theory or any other theory of parts. We indicate our Smith-candidate by pointing to it, rather than by saying what its parts are. This makes the crucial difference that the verb in the indeterminate sentence isn't "constitutes," which is vague, but the "is" of identity, which is precise.

The problem of the many requires that, for each ordinary name of an ordinary thing, there be candidates in great profusion for what the name refers to. Without some substantial metaphysics of temporal stages, the Smith example doesn't give us that. It gives us at least two candidates. A proponent of the single-candidate thesis might contend that there are only two plausible candidates for what I refer to when I point and say "that." If Smith were unmistakably alive, I would be referring to Smith, so what I say would be true. If Smith were undeniably dead, then I would be referring to Smith's body, and since Smith's body outlasts Smith, what I say wouldn't be true. As it is, however, my statement is ambiguous between the true proposition that Smith is Smith and the false proposition that Smith's body is Smith, so what I have said lacks a determinate truth value.

The Smith example doesn't directly refute the single-candidate hypothesis, but it does the next best thing. It sweeps away our motives for taking the hypothesis seriously in the first place. The primary appeal of the hypothesis is that it purports to enable us to hold on to the natural account of *de re* propositional attitude and indirect speech reports. A secondary motive it that it upholds the folk-semantic account of the workings of what we ordinarily regard as denoting proper names and successfully employed demonstratives. But in the Smith example, the hypothesis doesn't fulfill either promise.

Russell has given us only a single example, but a single clear example is all we ever really need, and it's hard to imagine a clearer example of a *de re* belief than my belief of the individual I'm pointing to that it's Ebenezer Wilkes Smith. The individual my belief is about is right before my eyes, and I'm pointing directly at it. No one who denies that there is an x such that I say that x is Ebenezer Wilkes Smith because I believe that x is Ebenezer Wilkes Smith can credibly claim to uphold our commonsense understanding of *de re* belief and speech reports. But if there is an x such that I have said that x is Ebenezer Wilkes Smith because I believe that x is Ebenezer Wilkes Smith, then it is sometimes correct to say of a thing that a speaker has ascribed an attribute to it, even though the speaker's speech act hasn't singled out a particular thing as his referent, and it is sometimes possible for a thinker to believe of a thing that it has a certain attribute even though the speaker's mental state

doesn't single out a particular individual as the object of his thought. The single-candidate theorist faces a dilemma. Either she denies the *de re* description of what I have said, in which case she forfeits her claim to be upholding our commonsense understanding of *de re* indirect speech reports, or she admits that sometimes a *de re* indirect speech report can be correct even though the speaker's statement doesn't have a uniquely determined referent. Similarly for beliefs. Either she denies the *de re* description of what I believed, in which case she abandons her claim to have captured our naive intuitions about *de re* mental attitude attributions, or she admits that a *de re* belief report can sometimes be correct even though there isn't any particular thing that the thinker has in mind. But once we allow that true *de re* indirect speech and propositional attitude reports that lack uniquely determined referents can be found at least occasionally, we have no compelling reason to presume that such reports cannot be found routinely. The promise of the single-candidate hypothesis was that it would let us uphold the thought that there can never be a true *de re* report without a uniquely determined object the report is about. Now that we see that this thought cannot be sustained even with the single-candidate hypothesis, the hypothesis loses its appeal.

The conclusion we are driven to is one we might also have reached, or so I believe, by pursuing the line of argument in Chapter Two of *Word and Object* [21]. Inscrutability of reference is a fact of life. Get used to it.

Acknowledgement. A version of this paper was read at the centenary conference on Russell's paradox in Munich, organized by Godehard Link and Ulrich Albert. Discussion there was quite valuable. I also enjoyed helpful conversations with Brian McLaughlin, Ólafur Jónsson, and Judith Thomson.

References

[1] Burgess, John A.: 1989. Vague identity: Evans misrepresented. *Analysis* 49: 112–119.

[2] Chang, Chen Chung, and H. Jerome Keisler: 1990. *Model Theory*, 3rd edition. Amsterdam: North-Holland.

[3] Evans, Gareth: 1978. Can there be vague objects? *Analysis* 38: 208. Reprinted in [9]: 317.

[4] Fine, Kit: 1975. Vagueness, truth, and logic. *Synthese* 30: 265–300. Reprinted in [9]: 119–50.

[5] Goodman, Nelson, and Henry Leonard: 1940. The calculus of individuals and its uses. *Journal of Symbolic Logic* 5: 45–55.

[6] Heck, Jr., Richard G.: 1993. A note on the logic of (higher-order) vagueness. *Analysis* 53: 201–208.

[7] Johnston, Mark: 1992. Constitution is not identity. *Mind* 101: 89–105.

[8] Jónsson, Ólafur P.: 2001. *Vague Objects*. Cambridge, MA: MIT doctoral dissertation.

[9] Keefe, Rosanna, and Peter Smith (eds.): 1996. *Vagueness: A Reader*. Cambridge, MA and London: MIT Press.

[10] Leśniewski, Stanisław: 1927–1931. O podstawach matematyki. *Przeglad Filozoficzny* (30: 164–206); (31: 261–291); (32: 60–101); (33: 77–105) and (34: 142–170). Abridged English translation: *Topoi* 2 (1989): 3–52.

[11] Lewis, David K.: 1970. General semantics. *Synthese* 22: 18–67. Reprinted in Lewis, David K.: 1983. *Philosophical Papers*, vol. 1. New York: Oxford University Press, 189–229.

[12] Lewis, David K.: 1986. *On the Plurality of Worlds*. Oxford and Cambridge, MA: Oxford University Press.

[13] Lewis, David K.: 1988. Vague identity: Evans misunderstood. *Analysis* 48: 128–130. Reprinted in [9]: 318–320.

[14] Lewis, David K.: 1993. Many but almost one. In: J. Bacon *et al*. (eds.), *Ontology, Causality, and Mind*, New York: Cambridge University Press, 23–38.

[15] McGee, Vann: 1991. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis: Hackett.

[16] McGee, Vann: 1997. Kilimanjaro. In: A. Kazmi (ed.), *Meaning and Reference*. Calgary: University of Calgary Press. *Canadian Journal of Philosophy*, supplementary vol. 25: 141–63.

[17] McGee, Vann, and Brian P. McLaughlin: 1995. Distinctions without a difference. *Southern Journal of Philosophy* 33 supp.: 203–252.

[18] McGee, Vann, and Brian P. McLaughlin: 2000. The lessons of the many. *Philosophical Topics*. 28: 129–151.

[19] Mehlberg, Henryk: 1958. *The Reach of Science*. Toronto: University of Toronto Press. Excerpt printed in [9]: 85–88.

[20] Quine, Willard V.: 1956. Quantifiers and propositional attitudes. *Journal of Philosophy* 53: 177–86. Reprinted in: Quine, Willard V.: 1966 *The Ways of Paradox*. New York: Random House, 183–94.

[21] Quine, Willard V.: 1960. *Word and Object*. Cambridge, MA: MIT Press.

[22] Russell, Bertrand: 1923. Vagueness. *Australasian Journal of Philosophy and Psychology* 1: 84–92. Reprinted in [9]: 61–68.

[23] Tarski, Alfred: 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Logica* 1: 261–405. English translation by J. H. Woodger in Tarski, Alfred: 1983. *Logic, Semantics, Metamathematics*, 2nd edition. Indianapolis: Hackett, 152–278.

[24] Thomson, Judith J.: 1998. The statue and the clay. *Noûs* 32: 149–173.

[25] Unger, Peter: 1979. I do not exist. In: G. MacDonald (ed.), *Perception and Identity*, Ithaca, NY: Cornell University Press, 235–51.

[26] Unger, Peter: 1980. The problem of the many. *Midwest Studies in Philosophy* 5: 411–67.

[27] Van Fraassen, Bas C.: 1966. Singular terms, truth value gaps, and free logic. *Journal of Philosophy* 63: 464–95.

[28] Van Inwagen, Peter: 1990. *Material Beings*. Ithaca, NY: Cornell.

[29] Wheeler, Samuel C.: 1979. On that which is not. *Synthese* 41: 155–173.

[30] Williamson, Timothy: 1994. *Vagueness*. London: Routledge.

[31] Zadeh, Lofti: 1975. Fuzzy sets and approximate reasoning. *Synthese* 30: 407–428.

Massachusetts Institute of Technology
Department of Linguistics & Philosophy
77 Massachusetts Ave., Room 32-D931
Cambridge, MA 02139
USA

E-mail: vmcgee@MIT.EDU

What Makes Expressions Meaningful? A Reflection on Contexts and Actions

Albert Visser

Abstract. This paper is about contexts and linguistic actions. Various intuitions are presented concerning contexts. I consider the idea that sentences are the appropriate contexts to make words meaningful. I argue against a version of the context principle loosely based on what Dummett calls a truistic interpretation of the slogan: *the sentence is the unit of meaning*. Finally, I draw attention to the existence of sub-sentential expressions that are used for global discourse actions.

Before we begin our banquet, I would like to say a few words. And here they are: Nitwit! Blubber! Oddment! Tweak! Thank you!

Albus Dumbledore

1. Introduction

In this paper I discuss the question: *what kind of thing makes an expression meaningful*? The leading idea is that expressions are made meaningful by being uttered in appropriate contexts. But, then, what are utterances and what are contexts?

In section 2, I introduce various intuitions concerning context and action. Some context principles are presented and a distinction is introduced between utterances and pre-utterances.

In section 3, I discuss the question whether the sentence has a primary role to play as meaning-conveying context. I argue against a version of Frege's context principle. The principle that I (try to) refute is loosely based on a supposed truism, discussed by Dummett in [3], to the effect that *the sentence is the unit of meaning*. I feel that a careful refutation of such principles is important, because adherence to such a principle may form an obstacle to a proper understanding of discourse and discourse actions. Moreover, forcing sentences into a role they are not meant to play, obscures the nature of sentences as units of local discourse organization.

Finally, in section 4, I argue that uses of sub-sentential expressions may have trans-sentential roles. This section is directly related to ideas from Dynamic Predicate Logic and Discourse Representation Theory. I will, however, give no technical realizations of the ideas sketched.

Conventions: I will use double quotes, italics or display when expressions are mentioned. I have no extra conventions for distinguishing, for instance, mention of a string of letters from mention of a word. In cases where it counts, I will simply use expressions like "the Dutch word "verflucht"", etc. I will use italics and display when text is quoted. Both italics and single quotes will be used for a variety of further purposes. Especially, the use of italics is somewhat overloaded. I hope that the context will resolve ambiguities in all cases.

2. Text in Context

In this section, I introduce various intuitions concerning *context*. I discuss, in passing, the notion of *utterance*.

2.1. Context & Language

Let us look at an exchange of letters between a Roman father and his son: *filius:* RUS EO. *pater:* I. (Allegedly, this is the shortest exchange of letters ever.) This translates into something like: *son: I'm going to the country. father: do go.* I don't know how many angels can dance on a needle's point, but surely lots of things can occur at the same place in a text. Apparently "I" is/carries: *an assertion/speech act, a text/discourse, a sentence, a word, a syllable, a morpheme, a letter, a visible item.*

Note that it is the Latin word "I" that is made present here and not the English word "I". Clearly, the physical item present on the page does not *intrinsically* carry all other items. Something extra is needed. This something is given with the *utterance situation*. I will call this something: $context_1$. Different items in the list demand different 'parts' or 'aspects' of the utterance situation for their individuation. For example, the German word "verflucht" (*damned*) is written in the same letters as the Dutch word "verflucht" (*smell of paint*). Yet, they are quite different words. Thus, if we compare a German (written) utterance of "verflucht" and a Dutch (written) utterance of "verflucht", I would be inclined to believe that the individual aspects of these utterances that are relevant for making the string of letters "verflucht" available are of the same type. However, the aspects that are relevant for making the words available cannot be of the same type.

$Context_1$s aren't always *individuating contexts*. E.g., a context provides the speaker, the time of speech, ... Thus, it contributes to the interpretation of the words "I" and "now". We can study such contexts taking the occurrence of the words for granted, as is done in the logical literature following Kaplan's pioneering paper, see [24]. Here are two basic intuitions.

1. $Context_1$ is co-individuated with the material it makes present.

2. Context$_1$ is of a different kind than the material it makes present.

Our first intuition has non-reductionist implications. Assuming that the material made available by a context is not by itself 'neutral', in the sense of unconnected to human intentionality, the context will not be neutral either. Thus, I do not think that, for instance, the existence of words has a neutral explanation, if only we consider enough of their 'environment'.

Our discussion to this point suggests some possible 'context principles'. For example,

α The fundamental form of language is text in context$_1$.

β An expression only exists in the context$_1$ associated with an utterance.

These proposals are more like schemata for principles than fully articulated principles, given the unclarity of such expressions as "text" and "utterance". This schematicity can have some advantages. E.g., we can understand *text* at different levels. Text might be strings of symbols, underlying graphs, strings of words, fully meaningful text, … . For each level of text the principle stipulates the need for an appropriate notion of context.

The advantage of principle (β) is that it explicitly relates contexts to intentional actions, viz., utterances. To give the principle as stated some plausibility, (i) we must construe the notion of utterance rather broadly. Moreover (ii), as I will show, utterances might be the wrong 'level' to serve for individuation of expressions.

Ad (i): we want to allow utterances that are not necessarily assertions, questions, etc. We want an address on an envelope or a list of items to be also utterances. "Salami" as occurring on a shopping list is surely a word. So, by (β), it must be part of an appropriate utterance.[1]

Ad (ii): there are problematic utterances, were no author of the utterance seems to be available. Consider the case of an ATM machine. I put my card in the machine and the machine asks: (1) *do you want to know your balance*? At a later time you put your card in the same machine and the machine asks: (2) *do you want to know your balance*? I submit that (1) and (2) are utterances and that they are different utterances. The utterance determines the addressee and the addressees of (1) and (2) differ. On the other hand, I do not think that the machine utters these utterances. The machine is lacking the right kind of intentionality to be the author of an utterance. Also it is not the programmers of the machine or the directors of the bank who utter these utterances, as it were indirectly. It would be miraculous if one could produce so many utterances without doing anything. Rather the programmers made a kind of pre-utterance that is activated by certain circumstances to produce/instantiate an utterance. Intentionality comes in at the level of the pre-utterance.

[1] What about the entries of a dictionary? Are they words? A dictionary lists words, not the items referred to by the words. We could simply take the entries of the dictionary to be *names* of the words listed and, hence, as words.

It is attractive to modify our principle and say that the words already exist at the level of the pre-utterance, even if in the case of the ATM example it is somewhat hard to say what precisely constitutes the pre-utterance. Consider a book in which the author addresses the reader: *dear reader* … . With respect to every reader, we seem to have a different utterance—even in the case where we consider subsequent readers of the same copy of the book. If we use the modified principle, we can simply say that the book contains an occurrence of the word "reader" at the given location. Under the original principle, we have to say either that in the book *an und für sich* on that location just a 'pre-word' exists that instantiates a word on any occasion of reading, or *alternatively* that, with any possible reader, a fixed corresponding utterance is associated—*a priori*, so to speak—and, hence that there are a large number of words, one for every possible reader, already present at that location in the book. I do not think either of these last two possibilities is particularly attractive. Thus, my preference would be to go for a modified version of (β). Perhaps a first attempt to define *text* could be: a text is an abstraction from a structured sequence of pre-utterances.

A further question is as follows. Do we find different pre-utterances at line n in different copies of the same book? This discussion is outside of the scope of this paper. My hunch is that they do contain the same pre-utterance and that, thus, the occurrences of the word "dear" on the given location in the various copies are strictly identical.[2]

In the next subsection, I consider Context in a more technical setting.

2.2. Contexts in Logic & Computer Science

Context_1 has an important analogue in Predicate Logic: the context of a formula is an assignment or a set of assignments. The view of assignments as contexts is exploited in Dynamic Predicate Logic where meanings are modeled as context modifiers. A formula, modeling an utterance, does not only demand an input assignment for its interpretation, the process of its interpretation also modifies the input assignment by creating and filling certain new variables / files / discourse referents. See [15], [23] and [32].

Our notion of $\mathit{Context}_1$ belongs to the (*context*/*text*) *duality*. In technical developments we also have a (*context*/*content*) *duality*. Let us call the associated notion: $\mathit{Context}_2$. The most familiar example of context_2 is the fibered view on Predicate Logic. Here the set of free variables is supplied locally with the formula. So the primary syntactic objects are of the form $\langle V, A\rangle$, where V is a (finite) set of variables and where A is a formula with free variables among the elements of V. The meaning of $\langle V, A\rangle$ over a model with domain D is of the form is $\langle V, F\rangle$, where $F \subseteq D^V$. In this framework the quantifiers appear as adjoints of the embedding functor:

$$\mathsf{emb}_{V,V\cup\{x\}} : \langle V, F\rangle \mapsto \langle V \cup \{x\}, \{g{\in}D^{V\cup\{x\}} \mid (g \restriction V) \in F\}\rangle.$$

[2] I think some serious rethinking of the familiar notions of *type*, *token* and *occurrence* is long overdue.

Another example of context$_2$ is formed by the type assignments to variables in Martin-Löf type theory.[3]

Finally, we have, in term rewriting, the notion of context as string / formula / term with a hole. Let us call this *Context*$_3$. (See chapter 2, by Jan Willem Klop and Roel de Vrijer of [40].)

It would be quite satisfying to see what—*vom höheren Standpunkt*—the connections between the various notions of context are.

3. Sentences as Contexts?

The most famous principle regarding contexts is Frege's Context Principle. Here is Frege's most well-known formulation of the principle.

> ▷ Nur im Zusammenhange eines Satzes bedeuten die Wörter etwas.
> See [11]: §62, 71.

The principle is often paraphrased as: *only in the context of a sentence does a word have meaning*. It is hard to see why such a special role should be given to sentences. Sentences, like words are syntactical items. Why should, with respect to meaning, one syntactical category be prior to another one? And what kind of priority are we talking about here?

In this section, I will critically discuss a version of the context principle derived from an interpretation, due to Dummett, of the slogan: *the sentence is the unit of meaning*. Our discussion will be structured around various remarks and ideas from Dummett's classical book *FREGE, Philosophy of Language*, or: FPL [3]. However, it is important to note that my explication of the principle goes beyond what is stated in Dummett's text.

Conventions: Sentences in their role as contexts seem to be a mixture of context$_1$ and context$_3$. In this section and the next, I will drop the subscripts. Contexts will be mostly contexts$_1$. Also I will ignore subtleties like the distinction between utterances and pre-utterances.

3.1. Acts and Moves in the Language-Game

Dummett makes a distinction between linguistic acts, i.e., acts constituting a full-blown move in the language-game and other acts playing a role in producing language. Here is an analogy. Checking what cards one has, is certainly an act, possibly even an essential act, in a card-game, but it does not constitute a move in the game. According

[3] For an example of an attempt to build a semantics by iterated use of context$_2$, see [43].

to Dummett, as I reconstruct him, uttering a non-sentential expression is clearly an act. It is just not a linguistic act, i.e., a move in the language-game.

One may have doubts about this distinction. After all, we were never provided with a list of the rules of the language-game. So, how do we know what does and what does not constitute a move in the bewildering variety of actions involved in language-use? In other words, the game metaphor *is* indeed just *that*: a metaphor. So, why should there be a definite notion about what constitutes a proper move?

In what follows, I will simply follow Dummett in the assumption that some clear-cut distinction can be made between linguistic acts or moves in the language-game on the one hand and other acts involved in the production of language. The assumption is a presupposition for Dummett's interpretation of the slogan: *the sentence is the unit of meaning*. If the assumption fails, the principle we are going to discuss, will be meaningless.

In this paper, I will use *utterance* in a broad sense: utterances are acts, but not necessarily moves in the language-game. For example, the utterance of a sentence will contain the utterances of various words as sub-utterances.

3.2. A Truism

According to Dummett, it is a truism …

> … that we cannot say anything by means of a sequence of words that stops short of being a sentence—cannot make an assertion, express a wish, ask a question, give a command, etc., in short do what Wittgenstein called 'making a move in the language-game'—except where the context supplies a supplementation of the words spoken that amounts to a sentence embodying them.

His reason is:

> … (in a logical rather than a typographical sense) an expression with which we can make a move in the language-game (or 'perform a linguistic act') is precisely what a sentence is. ([3]: 3)

Dummett views the slogan: *the sentence is the unit of meaning*, as a way of summarizing the supposed truism.

The true context principle is not this truism, according to Dummett. The true context principle provides the philosophical insight behind the truism. Dummett's formulation of the underlying insight is as follows. *In the order of* explanation *the sense of a sentence is primary, but in the order of* recognition *the sense of a word is primary* ([3]: 4). Our version of the context principle will not be Dummett's, but an expanded version of the supposed truism.

Note the clever exception that Dummett makes in his formulation of the supposed truism. This move is designed to block an almost inexhaustible list of counterexamples

against the principle. This exception will emerge as part (b) of our version of the sentential context principle.

3.3. A Truistic Context Principle

The interpretation of the slogan: *the sentence is the unit of meaning*, doesn't say anything about how arbitrary expressions derive their meanings from their role in sentences. What if an expression α occurs as part of a non-sentential expression β that amounts to a sentence γ due to contextual supplementation? Does α derive its meaning from γ? The Truistic Context Principle addresses such matters.

Let us say that a linguistic act, or move in the language-game, is *minimal* if it has no proper part, in the sense of sub-act, that is again a linguistic act. Let us call our version of the context principle TCP, for Truistic Context Principle. TCP consists of two theses.

a) An utterance of an expression is a meaningful use of that expression, if that utterance is part of an action that constitutes a minimal move in the language-game.

b) Consider a meaningful utterance u of an expression α. Let u be part of an action u' that constitutes a minimal move in the language game. Suppose u' takes place in a certain context c. Then either u' is an utterance of a sentence β or there is an alternative action v', in the same context c, that implements the same (or sufficiently similar) move in the language-game as u', such that v' is an utterance of a sentence γ and such that v' contains an utterance v of α that is sufficiently similar to u.

The qualification that 'v must be sufficiently similar to u' is designed to block examples in which α occurs in the expansion γ in a way not corresponding to the first use u of α. Suppose, e.g., someone says to his spouse with an angry face: *food*, and the expansion is: *I want my dinner now, for, as your husband, I have the right to get my food in time*.[4] (The claim to have a right to food is wholly conveyed by the angry look.) Then, probably, the use of *food* in the expansion is not sufficiently similar to the original use of the word. I am afraid that it would be no mean task to find a satisfactory explication of the notion of *sufficient similarity* as employed here.

Here is TCP in a different formulation. *Intra-sentential uses of non-sentential expressions derive their point from the use made of the embedding sentences to make a move in the language-game. Stand-alone extra-sentential uses of non-sentential expressions receive their point because these uses can be simulated by intra-sentential uses of the same expressions. Thus, there must be some appropriate relationship between the stand-alone utterance and the embedded utterance. It is not sufficient that*

[4] I neither do recommend such behaviour, nor support the attitudes underlying it.

the expression occurs in the sentence the use of which results in the same linguistic act as the use of the stand-alone expression in an appropriate context.

Note that (a) follows from two more attractive theses.

a1) An utterance of an expression is a meaningful use of that expression, if that utterance is part of an action that constitutes a move in the language-game

a2) Utterances with the part-of ordering form a well-founded partial ordering.

In the following subsections, I will discuss three issues concerning TCP. Some important issues will remain undiscussed. For example, not only in Dummett's philosophy, but in the mainstream logical tradition, a number of assumptions about sentence-structure and about the role this structure plays in the semantic evaluation process play a central role. I feel these assumptions need some substantial rethinking—if we are after a better understanding of *language*. This area of problems is of direct importance to the present discussion. In appendix A, I will make some sketchy remarks about these matters.

3.4. Existence of Expansions

For the moment, I will assume that sentences are sentences in the usual sense. In the next subsection 3.5, I will consider what happens, when we (re)interpret sentences according to Dummett's definition.

Is it always possible to find sentential expansions to accommodate examples of unembedded meaningful utterances of words, as promised in (b) of TCP? Let us consider some examples.

1. *With what shall I cut it? With an axe, dear Henry.* The answer, *With an axe, dear Henry*, can clearly be expanded to: *You shall cut it with an axe, dear Henry*. The expansion is rather canonical in such cases. This example illustrates that in some cases convincing expansions are available.
2. A child is walking with his mother. A cow is grazing in the field. The child says: *cow*. We could expand this to: *there is a cow*, or to: *that is a cow*. The child is doing many things at the same time: showing off its newly acquired linguistic competence, drawing attention to the cow, and so on. None of these is particularly well described as *asserting a truth* or *expressing a thought*. I submit that, e.g., *drawing attention to* should be considered as one of the primary kinds of a linguistic act. The example illustrates that the conventional repertoire of linguistic acts, *making an assertion, expressing a wish, asking a question, giving a command* ([3]: 3), should be extended to acts like *drawing attention to something*.
3. Consider an address on an envelope: *Harry Potter, 4 Privet Drive, Little Whinging*.

It would seem to me that an address is a meaningful use of language. Here is a possible expansion: *Dear mailman, please deliver this letter to Harry Potter. He lives in the Privet Drive, no. 4 in Little Whinging.* Well, that's two sentences. We can remedy it by inserting "and" or ";" between them. Still, it is doubtful whether the expansion gives us precisely the same linguistic act as simply the address. For instance, someone could meaningfully use an address without knowing that a mailman delivers the letters, even believing that letters are delivered by other means. Of course, we can find other expansions like: *deliver this letter to Harry Potter, 4 Privet Drive, Little Whinging.* This would make an address a command. How do we know that it is not simply the assertive statement: *This letter is intended for Harry Potter, 4 Privet Drive, Little Whinging*?
This example illustrates that it is probably impossible to find a sentential expansion that really simulates the linguistic act of writing an address. All expansions add too much.

4. Consider a shopping list: *2 juice, ….* A possible expansion is: *I want to buy: 2 juice, ….* Another: *whoever does the shopping has to buy: 2 juice, ….* Or perhaps: *Do get: 2 juice, ….*
This example illustrates the same problem as the previous one: expansions add too much. Moreover, I confess to the intuition that the understanding of a sentence containing a list partly rests upon our understanding of lists. Of course, a list should have a point. It is always a list with a purpose. This purpose is, in a sense, not part of the list. The question is: must this point be codified in a sentence?

5. Consider a title of a chapter, e.g., *De Lispeltuut*. Here is a possible expansion: *this chapter is about the Lispeltuut.* Another possibility would be *in this chapter I am going to tell you about the Lispeltuut*.
I guess these expansions are pretty close to what one wants to convey with the chapter title. However, if we write the expanding sentence as a chapter title, it is the place where it occurs—not its content—that makes it into a title. The point is that the linguistic act of giving a chapter title cannot be subsumed under *assertion*.

6. Consider counting the number of guests at a party: *1, 2, 3, ….* Here is an attempt to give an expansion: *There is at least one guest. There are at least two guests, There are at least three guests. ….* Again these are many sentences. We could remedy this by inserting *and* between the sentences. I have grave doubts that any such an expansion really is equivalent to counting. For example, from the current expansion it would follow that the discourse *1, 2* is logically equivalent to its sub-discourse *2*. Sounds funny.
This example illustrates that there is no plausible expansion for counting. Note that the example is particularly relevant since the status of number words was Frege's primary concern in Grundlagen.

7. I see my student Joost on the other side of the road. I call him: *Joost*.

It is hard to imagine an sentential expansion that doesn't add unwanted material.

8. Consider David Kaplan saying: *Albert, Albert, Albert*, at the end of my talk at the Russell conference.
   How could one ever find an expansion exactly capturing the impact of *that*?

In some of our examples (e.g., addresses) there are only expansions that implement a linguistic act that has the original act as a part, but such that every one of these expansions adds too much. One way to counter these examples would be to weaken thesis (b) in the appropriate way. Examples like counting seem to lack any appropriate expansion. The upshot is that there is at least reasonable doubt that all meaningful uses of words can be expanded faithfully to sentences, i.e., in such a way that, keeping the original context of utterance fixed, the original linguistic act is simulated by the new one.

3.5. Redefining the Sentence

Dummett's reason for thinking that the supposed truism is indeed a truism, rests on his definition of sentence. I will scrutinize this definition in more detail to see whether it can serve to make TCP more evident.

On page 34 of FPL, Dummett states his definitions of sentence and word as follows.

> A sentence is a linguistic unit: it is the smallest bit of language which one can use to *say* anything—to 'make a move in the language-game', in Wittgenstein's terminology. A word, regarded from a different viewpoint, is another kind of linguistic unit: it is the smallest bit of language to which one can attribute a sense.

Evidently 'smallest' cannot mean smallest qua number of letters or phonemes in the embodiment of the expressions: this would lead to absurd results. A first guess is that it means: minimal in the constituent ordering (with the given property). For instance, the definition of sentence will become:

- ▹ Something is a sentence if it can be used to make a move in the language-game and if it has no proper constituent that can be used to make a move in the language game.

Under this interpretation, *words* and *sentences* become a sort of atoms. Indeed, Dummett remarks in a footnote that his definition of word *will not always coincide with what the printer makes*, mentioning, e.g., verb-endings as an example. Thus, Dummett is aware that his 'words' are more like morphemes.[5] Words in the usual sense become complex expressions. For instance, "unutterable" is an expression with three constituent words: "un", "utter" and "able".

[5] See [26], for an introduction.

Now consider (1) "John came and Mary went" and (2) "John said that Mary went". Here both (1) and (2) have the smaller constituent "Mary went", so they are not sentences. So it would seem that the word "and" as a connective of sentences and the word "that" in its role of introducing a sentential clause, never occur in a sentence. Hence, according to the Dummett-Frege philosophy, they do not have meaning.

Of course, this is not what Dummett intends. I think that the right reconstruction is in the spirit of the formulation of our thesis (a):

- ▷ Something is a sentence if it can be used to make a minimal move in the language-game. I.o.w., an expression α is a sentence if there is an utterance u of α such that u constitutes a minimal move in the language-game.

E.g., when we say "John said that Mary went" no move in the language-game is supposed to correspond with the sub-utterance of "Mary went", which does not imply that there are no *unrelated* utterances of "Mary went" that do constitute moves in the language-game. Under this reconstruction Dummett simply misstated his definition: he married the 'smallest' to the wrong partner. He should not have said: *it is the smallest bit of language which one can use to say anything*, but: *it is the bit of language that can correspond to a smallest way of saying something*.[6,7]

The principle TCP clearly does not follow from Dummett's definition of sentence, since we cannot infer that every minimal move in the language-game (that involves the use at least one expression) can also be made by an utterance of a sentence (containing that expression in the right way)—as (b) implies. We need something like the assumption that, *every linguistic act can be made as the utterance of some expression*. (In fact, even this assumption is not enough, since it does not guarantee the possibility of 'simulation'.)

How does Dummett's definition of sentence interact with the clever phrase: *except where the context supplies a supplementation of the words spoken that amounts to a sentence embodying them*? In the phrase it is explicitly conceded that non-sentential expressions, in certain contexts, can be used to the same effect as an utterance of an appropriate sentence. So, in these contexts, these expressions are used with a linguistic act as result. But if that is so, they seem to be, by Dummett's own definition, themselves sentences and not in need of further supplementation. So we might conclude that (i) many more things are sentences than one would naively suspect and (ii) thesis (b) of TCP is superfluous, since it is automatically fulfilled: in all cases we only need the identity-expansion.

Is it possible to block this proliferation of 'sentences'? One might argue that, *yes*, there are actions that amount to moves in the language-game, such that (i) the only

[6] If we compare the discourses "John came and Mary went" and "John came. Mary went", then we may feel some doubt even about the amended definition. If we assert "John came and Mary went", do we then not, ipso facto, assert "John came"?

[7] There is a different line of questioning Dummett's definition of sentence. One can ask oneself what precisely a 'bit of language' is. Is it a neutral thing, like an equivalence class of equiform tokens (as Tarski has it in his footnote 5 of [39])? Or is it something different? I will not pursue this line in the present paper.

linguistic aspects of these actions are the utterance of an expression α, but (ii) *no*, these actions are not themselves utterances of α. So, for instance, the move in the language-game made by saying "with an axe, dear Henry" is not implemented by the utterance: *with an axe, dear Henry*, but by a larger action, presumably containing 'hidden parts'.

Maybe the defect of the utterance of "with an axe, dear Henry" is that the non-sentential expression exploits the context in a more substantial way than the expanding sentence. However, that difference cannot be simply that sentences do not need a context. Consider "he was quite happy". Clearly, context is needed in an essential way to provide the reference of "he" and the time corresponding to "was". Even if we would allow only eternal sentences, some context is needed to make an act an utterance of a given eternal sentence.

I suspect that no good account of *linguistic act* is possible, to obtain, via Dummett's definition, a class of sentences that even roughly coincides with the class of things usually called sentences. To that purpose, you simply cannot avoid things like saying that a sentence is organized around the verb, that around it the arguments are specified according to various strategies, etcetera. On the other hand, simply coining a new word "sentence" to replace the old one, makes the slogan: *the sentence is the unit of meaning* not just empty, but also misleading. Nothing about the form of the words is relevant, just the underlying move in the language-game. The slogan should have been: *the move in the language-game is the unit of meaning*.

3.6. Is TCP Strong Enough?

In the previous subsection, we formulated a doubt about whether TCP really supports the slogan that the sentence is the unit of meaning in any intuitive sense. It seems to me that this doubt even exists independently of Dummett's definition of sentence.

In TCP, it is the fact that the word is part, in an appropriate way, of a linguistic act that *makes* the word meaningful, not the act-preserving sentence that we construct around it. Only if you could claim a more substantial status for the expansion than just its possibility in principle, would the expansion really contribute to the possession of meaning by an uttered word. Perhaps, always when we utter a stand-alone non-sentential expression, we actually have in mind an expanding sentence. From her side the hearer only understands the utterance of the stand-alone non-sentential expression if she infers the sentence that the speaker had in mind, or a close equivalent, from the utterance plus the associated context. However, all such theories seem to be quite implausible.

3.7. Conclusion

I argued against one version of the context principle: TCP. Regrettably, many variants escape these criticisms. We could, for example, grant that there are irreducible extra-

sentential uses of phrases that amount to full-blown linguistic acts, but still claim that the understanding of these uses is derivative from our understanding of uses of the same words in sentences.[8] Of course, I am committed to the thesis that there is something generally wrong about sentential context principles.[9]

In the next section, I will follow a different line. Not stand-alone uses of non-sentential expressions but certain intra-sentential uses of non-sentential expressions will be considered. I will claim that these uses implement acts whose effects are not restricted to the sentence in which they occur. The effects are trans-sentential. I think that examples along this line are more effective to show *that the intuition behind the sentence-as-unit-of-meaning idea is essentially misguided.*

4. Discourse

In this section, I will discuss examples like "a man" and "he". I argue that the understanding of these phrases involves understanding their use in trans-sentential discourse. Understanding the indefinite article "a" is to understand its use in expanding open-ended discourse, not its use in a single sentence. The actions performed by uttering such phrases can be clearly formulated. Indeed, we are close—in dynamic semantics—to giving satisfactory formal modelings of these actions. Moreover, the acts performed by uttering these phrases serve the purpose of facilitating the transmission of information—surely one of the major aims of language. I submit that these actions have as good a claim as anything to embodying full-blown, honest-to-god moves in the language-game.

4.1. Referring Expressions in Discourse

Frege, as reconstructed by Dummett, holds that the sentence is the primary linguistic entity to perform linguistic acts. Larger discourses also embody linguistic acts in virtue of forming a sequence of sentences.[10] However, uses of intra-sentential non-sentential expressions do not constitute full-blown linguistic acts or moves in the language game.

Let us consider the case of saying a name in discourse. Robert Sokolowski discusses in his paper [37] (reprinted as chapter 9 of [38]) the case where he is dealing

[8] Note, however, that such weakenings of TCP, strengthen the considerations of subsection 3.6.

[9] I even think that some related non-sentential context principles are false. Consider, e.g., the alternative context principle: *an utterance of an expression is a meaningful use of that expression, if that utterance is part of an utterance that expresses a thought.* There are aspects of the meaning of expressions that cannot be reduced to any specific thought they help to express.

[10] *... to a first approximation, the understanding of speech just is the understanding, in succession, of the sentences composing it* ([4]: 373). In a footnote to the same page: *A second approximation would have to take account of one inessential feature, the back-reference of demonstratives, and one essential one, the genuine or purported logical connections between the assertions made.*

with 'Topper'. Now Margaret says: *Topper*. What Margaret does is, according to Sokolowski, *targeting Topper for disclosure* and *moving Topper to presentational form*. Something like that seems to be more or less right. Consider for example a half-heard discourse.

> … Topper mumble mumble he has already eaten mumble …

The embedding sentence in which "Topper" first occurs is irrelevant for the long range discourse effect of the utterance of "Topper". Uttering "Topper" (in the right kind of context) contributes to the thoughts expressed by the utterances of sentences in which "Topper" does not occur. Saying "Topper" makes him salient to be picked up later by an anaphor. In some versions of Discourse Representation Theory (DRT), see [22], this idea is reflected by analyzing saying "Topper" as addition to the current information of the following discourse representation structure:

x
$Topper = x$

The idea is that saying "Topper" implements the following action: a new discourse referent x is introduced to be picked up by subsequent uses of anaphors—this is witnessed by the occurrence of x in the top box—, and that simultaneously the value of this referent is constrained to be the value of "Topper"—witnessed by the identity in the lower box, see [22]. I do not think this is completely right yet. The local role in the embedding sentence of saying "Topper" is not well represented. The right analysis must be something like this: saying *Topper* in a sentential context simultaneously realizes the following discourse actions.

1. Topper is made salient. A global referent is introduced to be picked up by uses of anaphors in subsequent discourse.

2. The value of this referent is constrained to be the value of the global referent *Topper*.

3. The value of the current argument role (e.g., *subject*) is identified with the value of "Topper". Topper is *made* the subject of the sentence uttered.

Similarly, saying *a man* in an appropriate context amounts to the instructions (i) introduce a new discourse referent, (ii) constrain the possible values of this referent to be men, and (iii) set the value of the new referent equal to the value of the current argument role.

I turn to the case of definite descriptions. It seems to me that the primary role of definite descriptions is as anaphors. The discourse actions corresponding to uttering a definite description are like this:

1. Import an appropriate discourse referent from the context. The material in the 'body' of the description is there to help locate the intended referent.

2. Make the imported referent salient in order to pick it up again by later uses of anaphors.

3. Make the value of the imported referent equal to the value of the current argument role (like *subject* or *object*).

Note the differences with Russell's great theory of definite descriptions, see [34, 35]. (i) Our descriptions are terms. (ii) The 'identifying' part is active at another level: we want to select a discourse referent out of a stock of contextually given discourse referents, not find a uniquely determined denotation.[11] Of course, information about the value of the discourse referent is ipso facto information about it.

Uses of names, indefinite descriptions and definite descriptions implement transsentential structuring discourse actions. In uttering a sentence that contains such a phrase one not only makes an assertion, but one also makes certain referents available for further use in the discourse.

Such discourse actions often have two kinds of incompleteness. (i) They call upon the context for certain material to use and (ii) they create an expectation of further actions (to give them a point). If I simply say *a man* followed by silence, this would seem surpassingly strange. Does this kind of incompleteness disqualify these actions for the desired status of move in the language-game? I do not think so. Let us quote Dummett again:

> Uttering an expression which refers to an object has in itself no significance considered in isolation from any context which determines what the person is trying to do by uttering that expression; only a context will give a point to the utterance. If, e.g., I say, out of the blue, 'the highest mountain in the world', I have indeed uttered an expression which is endowed with a certain sense, and thereby a certain reference; but, until I have indicated what point I had in mind in doing this, I cannot be considered to have done anything *right* or *wrong*. The natural reaction to my utterance would be, 'What about it?', or, 'Well, go on.' … ([3]: 297f.)

I couldn't agree with him more. However, then he[12] goes on to assign sentences a special status.

> … Now the utterance of a sentence does not require a special context to give it point, but is governed by a general convention—at least in certain types of situation—that in uttering them we are understood as saying that their reference is truth.

[11] Donnellan's *the man over there drinking champagne* clearly is sufficient to pick up the discourse referent, even if the embedded information is false for the denotation. Here is another suggestive example: *A man and his son entered the room. The son greeted everybody loudly. Together with them a second of the man's sons came in. He was remarkably silent.*

[12] Or, more precisely, Frege as reconstructed by Dummett.

I think this just isn't so (even after suitably adapting the typically Fregean bit about 'their reference being truth'). Utterances of sentences may exhibit precisely the same kind of incompleteness of standing in need of a point to be supplied by subsequent discourse. If I am reporting what happened at a party and if I say (i) *A man came in*, then further disclosure would be expected to give the utterance a point. Similarly, when in mathematical discourse, I say (ii) *Let x be a prime number greater than 7*, I would be expected to go on. This expectation-generating aspect is no reason at all to deny that what one does in saying (i) or (ii), in the right kind of context, amounts to a full-blown move in the language-game. So why could saying, in the right kind of context, *the highest mountain in the world* not also amount to a move?

Dummett contrasts Frege's doctrine of the priority of the sentence, to Aristotle's view that the role of words is to express ideas and that the role of complex expressions is to express complex ideas ([3]: 3). The kind of view I propose here also diverges from Aristotle: saying a word can be a linguistic act, an act that cannot be simply identified with just the act of expressing an idea.

In appendix A, I discuss the relationship of the ideas described above with predicate logic. I also sketch a few ideas about discourse structure.

4.2. So What Is a Sentence?

At the end of the paper, it seems appropriate, to state my view of what the sentence is. In slogan: the sentence is the playground of a local strategy to get information into place. Sentence structure in a wide sense embodies directions to handle information. The sentence is centered around the verb with its associated argument structure. Sentences in different languages implement various strategies to connect referring phrases to the various argument roles.[13]

The sentence is not the only linguistic object embodying some strategy to get information into place. A list, for example embodies another strategy, but lists are not sentences and sentences are not lists.[14]

The nature of the sentence as the realization of an organizational idea is obscured by theories that make sentences the *sine qua non* of meaningful thought.

A. Sentence Structure & Discourse Structure

The presence of long range discourse actions distinguishes natural language from predicate logic. The way predicate logic is set up mirrors its original design ideas. We wanted a language that would allow *full and fully explicit control*. The use of

[13] I do not believe that *all* aspects of sentence structure are functional for directing information into place. But I do believe that some aspects are thus functional.

[14] For some remarks on sentence structure versus discourse structure, see appendix A.

explicit variable names, the specific scoping mechanisms that restrict the scope of a quantifier to the part of the parse tree below it, they are all embodiments of the desire for control. The various mechanisms for control make predicate logic quite different from natural language. As a corollary one must be very careful to project properties of predicate logic, that well-understood model, onto natural language. It is certainly quite wrong to think of predicate logic as revealing the underlying form of natural language, whatever that may be, in any direct or simple way.

One reason why the study of traditional predicate logic is misleading—if we want to reflect on the nature of language and meaning—, is the fact that it embodies a strict separation between the syntax of formulas and the syntax of proofs. Proofs are built on top of formulas. Such discourse actions as *suppose* and *unsuppose* in hypothetical reasoning are not part of the formulas, but part of the syntax of proofs.

The example of *suppose* and *unsuppose* as we would use it in (semi-)natural language illustrates an important theme: the analogy between some discourse actions and brackets. The 'discourse brackets' *suppose* and *unsuppose* enclose a stretch of discourse in which a certain assumption is entertained.[15] A similar view can be taken for *a man*. Saying *a man* opens a 'discourse mode' in which a new discourse referent has a value. (See [41] and [20] for a version of dynamic semantics, where this view becomes evident.)

Our examples suggest that the traditional separation between *categorematical* and *syncategorematical*, is not as clear-cut as one might think. A bracket can be alternatively viewed as corresponding to an action, and thus as a meaningful piece of text.[16]

In the classical view, we only understand what a sentence means after (i) it has been parsed, (ii) the meanings of all constituents are computed inside out, following the construction of the sentence. The same kind of view would be totally implausible for discourse structure. We cannot wait till a discourse is finished, before understanding may happen.

I propose that we view sentence structure and discourse structure as different but related ways to get information into place. Sentences are a local way of information management, which exploits the fact that the verb naturally comes equipped with 'argument places'. Discourse structure is a global way. The actions we associated with uses of referring expressions in subsection 4.1, have a portmanteau character: they split into parallel sub-actions. One is concerned with sentence structure, one is concerned with discourse structure and one is concerned with content. Thus, sentence structuring and discourse structuring run in parallel in natural language. We also increase the content. However, there is no reason to link the content increasing actions in general more to sentence structure than to discourse structure. For instance, if we say *a man*, we constrain the value of the discourse referent we just introduced to be a man. This constraint has global significance.

[15] A formal study of *suppose* and *unsuppose* as discourse brackets/actions was undertaken in [44].

[16] There is an analogy between the 'ambiguity' of brackets and the two ways of viewing Polish notation. See [42].

Acknowledgement. I thank Menno Lievers for many pointers to the literature and for his careful reading of the penultimate draft. (Unfortunately, I had no space to address *all* his criticisms.) I am grateful to Godehard Link, both for his comments and for his insistence, without which this paper would never have been finished. I thank the audience of my talk at the Russell conference for helpful remarks and comments.

References

[1] Aristotle: 1966. *Categories, and De Interpretatione*. Translated with notes by J. L. Ackrill. Oxford: Clarendon Press.

[2] Sciullo, Anna-Maria: 1988. *On the Definition of Word*. Cambridge: MIT Press.

[3] Dummett, Michael: 1981. *FREGE, Philosophy of Language*. London: Duckworth.

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

[5] Dummett, Michael: 1995. The context principle: centre of Frege's philosophy. In: I. Max and W. Stelzner, *Logik und Mathematik*, Berlin: Walter de Gruyter, 3–19.

[6] Frege, Gottlob: 1970. *Translations from the Philosophical Papers*. London: Blackwell.

[7] Frege, Gottlob: 1975. Über Sinn und Bedeutung. In [8]: 40–65. Also in [16]: 142–160. English translation in [6]: 56–78.

[8] Frege, Gottlob: 1975. *Funktion, Begriff, Bedeutung*. Göttingen: Vandenhoeck & Ruprecht.

[9] Frege, Gottlob: 1976. Der Gedanke. In: G. Patzig (ed.), *Logische Untersuchungen*. Göttingen: Vandenhoeck & Ruprecht, 30–53.

[10] Frege, Gottlob: 1983. Logische Mängel in der Mathematik. In: H. Hermes *et al*. (eds.), *Gottlob Frege. Nachgelassene Schriften*, Hamburg: Felix Meiner Verlag, 171–181.

[11] Frege, Gottlob: 1988. *Die Grundlagen der Arithmetik*. Hamburg: Felix Meiner Verlag.

[12] Frege, Gottlob: 1998. *Grundgesetze der Arithmetik I/II*. Hildesheim: Georg Olms Verlag.

[13] Frege, Gottlob: 1998. Begriffsschrift. In: I. Angelelli, *Begriffsschrift und andere Aufsätze*, Hildesheim: Georg Olms Verlag, 1–88.

[14] Greenberg, Marvin J.: 1996. *Euclidean and Non Euclidean Geometries*. Freeman, 3d edition.

[15] Groenendijk, Jeroen and Martin Stockhof: 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14: 39–100.

[16] Harnish, Robert M.: 1994. *Basic Topics in the Philosophy of Language*. New York: Harvester Wheatsheaf.

[17] Hilbert, David and Paul Bernays: 1939. *Grundlagen der Mathematik II*. Berlin: Springer.

[18] Hilbert, David and S. Cohn-Vossen: 1973. *Anschauliche Geometrie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

[19] Hofstadter, Daniel. R.: 1979. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books, Inc.

[20] Hollenberg, Marco and Cees Vermeulen: 1996. Counting variables in a dynamic setting. *Journal of Logic and Computation vol. 6* 5: 725–744.

[21] Kamp, Hans: 1981. Truth, interpretation and information. In: J. Groenendijk *et al.* (eds.), *A Theory of Truth and Semantic Representation*, Dordrecht: Foris, 1–41.

[22] Kamp, Hans and Uwe Reyle: 1993. *From Discourse to Logic I/II*. Dordrecht: Kluwer.

[23] Kamp, Hans and Jan van Eijck: 1997. Representing discourse in context. In: J. van Benthem and A. ter Meulen, *Handbook of Logic and Language*, Amsterdam & Cambridge: Elsevier & MIT Press, 179–237.

[24] Kaplan, David: 1989. Demonstratives. In: J. Almog *et al.* (eds.), *Themes from Kaplan*, New York: Oxford University Press, 481–563.

[25] Kaplan, David: 1990. Words. *Aristotelian Society* Supp. 64: 93–119.

[26] Katamba, Francis: 1994. *English Words*. London: Routledge.

[27] Lyons, John: 1995. *Linguistic Semantics: an Introduction*. Cambridge: Cambridge University Press.

[28] Manguel, Alberto: 1996. *A History of Reading*. New York: Penguin Putnam Inc.

[29] Mill, John S.: 1843. *A System of Logic*.

[30] Mill, John S.: 1994. *Of Names*. In [16]: 130–141.

[31] Montague, Richard: 1970. Universal grammar. *Theoria* 36: 373–398.

[32] Muskens, Reinhard, Johan van Benthem, and Albert Visser: 1997. Dynamics. In: J. van Benthem and A. ter Meulen, *Handbook of Logic and Language*, Amsterdam & Cambridge: Elsevier & MIT Cambridge, 587–648.

[33] Peirce, Charles S.: 1906. Prolegomena to an apology for pragmaticism. In: C. Hartshorne and P. Weiss, *Collected Papers of Charles Sanders Peirce*, Cambridge: Belknap Press of Harvard University Press.

[34] Russel, Bertrand: 1905. On denoting. *Mind* 14: 479–493.

[35] Russell, Bertrand: 1919. *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.

[36] Russell, Bertrand: 1940. *An Enquiry into Meaning and Truth*. London: George Allen and Unwin.

[37] Sokolowski, Robert: 1988. Referring. *The Review of Metaphysics* 42: 27–49.

[38] Sokolowski, Robert: 1992. *Pictures, Quotations and Distinctions*. Notre Dame: University of Notre Dame Press.

[39] Tarski, Alfred: 1944. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4: 341–375.

[40] Terese: 2003. *Term Rewriting Systems*. Cambridge: Cambridge University Press.

[41] Vermeulen, Cees F. M.: 1993. Sequence semantics for dynamic predicate logic. *Journal of Logic, Language and Information* 2: 217–254.

[42] Visser, Albert: 2001. *On the Ambiguation of Polish Notation*. Artificial Intelligence Preprint Series 26, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht,
http://preprints.phil.uu.nl/aips/.

[43] Visser, Albert and Cees Vermeulen: 1996. Dynamic bracketing and discourse representation. *Notre Dame Journal of Formal Logic* 37: 321–365.

[44] Zeinstra, Liesbeth: 1990. *Reasoning as Discourse*. Master's thesis, Department of Philosophy, Utrecht University.

Universiteit Utrecht
Heidelberglaan 8
3584 CS Utrecht
Netherlands

E-mail: Albert.Visser@phil.uu.nl

List of Contributors

John L. Bell is Professor of Philosophy and Adjunct Professor, Department of Mathematics, at the University of Western Ontario, London, Ontario, Canada.

Ulrich Blau is Professor of Logic and Philosophy of Science at the University of Munich, Germany.

Andrea Cantini is Professor of Logic at the Università degli Studi di Firenze, Firenze, Italy.

Solomon Feferman is Professor of Mathematics and Philosophy at Stanford University, Stanford, CA, USA.

Hartry Field is Professor of Philosophy at New York University, New York, NY, USA.

Harvey Friedman is Professor of Mathematics at Ohio State University, Columbus, OH, USA.

Sy D. Friedman is Professor of Mathematical Logic at the University of Vienna, Austria and Professor of Mathematics at the Massachusetts Institute of Technology, Cambridge, MA, USA.

Nicholas Griffin is Canada Research Chair in Philosophy and Director of the Bertrand Russell Research Centre at McMaster University, Hamilton, Ontario, Canada.

Kai Hauser is Professor of Mathematics at the Technische Universität Berlin, Germany.

Allen P. Hazen is Lecturer of Philosophy at the Unversity of Melbourne, Australia.

Geoffrey Hellman is Professor of Philosophy at the University of Minnesota, Minneapolis, USA.

Thomas Hürter is Editor, MIT Technology Review, Hannover, Germany.

Andrew D. Irvine is Professor of Philosophy at the University of British Columbia, Vancouver, Canada.

Gerhard Jäger is Professor of Theoretical Computer Science and Logic at the University of Bern, Switzerland.

Gregory Landini is Professor of Philosophy at the University of Iowa, Iowa City, IA, USA.

Shaughan Lavine is Professor of Philosophy at the University of Arizona, Tucson, AZ, USA.

Godehard Link is Professor of Logic and Philosophy of Science at the University of Munich, Germany.

Bernard Linsky is Professor of Philosophy at the University of Alberta, Edmonton, Alberta, Canada.

David Charles McCarty is Professor of Philosophy at Indiana University, Bloomington, IN, USA.

Vann McGee is Professor of Philosophy at the Massachusetts Institute of Technology, Cambridge, MA, USA.

Jan Mycielski is Professor of Mathematics at the University of Colorado, Boulder, CO, USA.

Karl-Georg Niebergall is Privatdozent of Philosophy, Logic and Philosophy of Science at the University of Munich, Germany.

Volker Peckhaus is Professor of Philosophy of Science and Technology at the University of Paderborn, Germany.

Dieter Probst is Research Associate in Theoretical Computer Science and Logic at the University of Bern, Switzerland.

Michael Rathjen is Professor of Mathematics at the University of Leeds, England, and at Ohio State University, Columbus, OH, USA.

Francisco Rodríguez-Consuegra is Professor of Philosophy at the University of Valencia, Spain.

Philippe de Rouilhan is Director of Research at the Centre National de la Recherche Scientifique (CNRS), Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST), Paris, France.

Peter Schuster is Privatdozent of Mathematics at the University of Munich, Germany.

Helmut Schwichtenberg is Professor of Mathematics at the University of Munich, Germany.

Holger Sturm is Researcher in Logic and Philosophy at the University of Konstanz, Germany.

Robert S. D. Thomas is Professor of Mathematics, University of Manitoba, Winnipeg, Canada.

Albert Visser is Professor of Philosophy at the University of Utrecht, The Netherlands.

Kai F. Wehmeier is Assistant Professor of Logic and Philosophy of Science at the University of California, Irvine, CA, USA.

W. Hugh Woodin is Professor of Mathematics at the University of California at Berkeley, Berkeley CA, USA.

Name Index

Ackermann, Wilhelm, 147n
Aczel, Peter, 149, 194, 195, 232n, 261
Aquinas, 192
Aristotle, 166, 192
Armstrong, David M., 592, 603, 606, 607
Ayer, Alfred J., 482
Azzouni, Jody, 581

Balaguer, Mark, 551, 584
Baltag, Alexandru, 149n
Barendregt, Henk, 236n
Barwise, Jon, 149n, 194
Bealer, George, 600, 604
Bell, John L., 20, 568, 574, 587
Benacerraf, Paul, 572
Berkeley, George, 497
Berkowski, Hermann, 511
Bernays, Paul, 142n, 193, 313
Bishop, Errett, 195, 227, 233n, 449
Blackwell, Ken, 495
Blass, Andreas, 147n
Blau, Ulrich, 21
Blumenthal, Otto, 504
Boole, George, 507, 509
Boolos, George, 222, 248n, 563
Bourbaki, Nicolas, 565
Bradley, Francis H., 3n, 353, 427, 432
Bricmont, Jean, 546
Bridges, Douglas, 242
Broad, Charles D., 422n, 487
Brouwer, Luitzen, 449, 451, 518ff
Brown, James Robert, 584
Bunge, Mario, 581
Bunn, Robert, 352
Burali-Forti, Cesare, 3n, 350n, 351
Burgess, John, 248n, 253, 473, 570
Buridan, John, 464

Cantini, Andrea, 20, 201
Cantor, Georg, 3n, 5, 93, 96, 250, 375, 534, 551, 553
Carnap, Rudolf, 14, 439, 442, 449, 523
Church, Alonzo, 12n, 13, 435n, 453
Chwistek, Leon, 442n, 446n
Cocchiarella, Nino, 389n, 595
Coffa, Alberto, 350
Cohen, Paul, 29, 33, 85, 93, 101
Conway, John, 311, 321
Crimmins, Mark, 581
Cusanus, Nicolaus, 96n

Davoren, Jen M., 22, 452
De Morgan, Augustus, 577
Dedekind, Richard, 350, 375, 412, 511, 553, 562, 569
Descartes, René, , 96n, 112, 497
Donnellan, Keith, 639n
du Bois-Reymond, Emil, 22, 523, 529
du Bois-Reymond, Paul, 22, 522ff
Dummett, Michael, 248, 630

Eames, Elizabeth, 496, 585
Ernest, Paul, 578
Etchemendy, John, 601
Euclid, 102
Evans, Gareth, 580, 616

Fechner, Gustav, 529
Feferman, Solomon, 12n, 13f, 19, 148n, 154n, 161f, 193, 200, 268n, 272, 594
Fernando, Tim, 150
Ferreira, Fernando, 251
Feyerabend, Paul, 17n
Field, Hartry, 15, 20, 584
Fine, Kit, 10n, 248n
Finsler, Paul, 193

Fleischer, Heinrich, 508
Frege, Gottlob, 2, 5, 9f, 11n, 20, 247ff, 350, 375, 412, 507ff, 534, 561, 579, 625
Friedman, Harvey M., 18
Friedman, Sy D., 18
Fries, Jacob F., 510

Galilei, Galileo, 4, 16
Garciadiego, Alejandro, 350
Gentzen, Gerhard, 455
Geuvers, Herman, 236n
Gödel, Kurt, 2, 6f, 12f, 18, 85, 93, 97, 98, 101, 142n, 146n, 449, 587
Goesch, Heinrich, 511f
Goldfarb, Warren, 252
Goodman, Nelson, 15, 449, 617
Grattan-Guinness, Ivor, 350
Grelling, Kurt, 510n, 511f
Griffin, Nicholas, 2n, 21, 489
Grim, Peter, 261
Grishin, V. N., 275
Grothendieck, Alexander, 148
Gupta, Anil, 595

Hacker, Peter M. S., 496
Hale, Robert, 248n
Hauser, Kai, 18
Hazen, Allen P., 10n, 22, 452
Heck, Richard, 248n, 251
Hellinger, Ernst, 506
Hellman, Geoffrey, 23
Helmholtz, Hermann von, 529
Hempel, Carl G., 496, 497
Hersh, Reuben, 578
Herzberger, Hans, 595
Hessenberg, Gerhard, 5n, 510f
Hilbert, David, 2f, 93, 102, 159f, 162, 351, 455, 517, 533, 551
Hodes, Harold, 581
Hürter, Tobias, 19
Husserl, Edmund, 18, 104, 110, 507f

Hylton, Peter, 393n, 489, 490, 496

Inwagen, Peter van, 581
Irvine, Andrew D., 15, 22
Isaacson, Daniel, 587

Jäger, Gerhard, 14, 18
Jané, Ignacio, 586
Jensen, Ronald, 86, 149f
Jónsson, Ólafur, 617, 621
Jordan, Camille, 612
Jubien, Michael, 592

Kaiser, Karl, 510
Kant, Immanuel, 4, 102
Kaplan, David, 395n
Keränen, Jukka, 570
Klaua, Dietrich, 321n
Kleene, Stephen, 288n
Klein, Felix, 5
Kneser, Martin, 236
Kotarbiński, Tadeusz, 537
Kreisel, Georg, 200, 585n
Kripke, Saul, 193, 288, 317n
Kunen, Kenneth, 87, 534

Landini, Gregory, 8n, 11n, 21f, 439, 442n, 452, 472, 476
Lavine, Shaughan, 15, 19, 22, 153, 163, 165f, 173, 177n
Leblanc, Hugues, 462
Lehman, Hugh, 586
Leibniz, Gottfried, 96n, 484, 489
Leonard, Henry, 617
Leśniewski, Stanisław, 165n, 617
Levy, Azriel, 141
Lewis, David, 193, 600, 601, 613n, 619
Lindström, Ingrid, 202
Link, Godehard, 136n, 138n, 140n, 150, 183n
Linsky, Bernard, 21, 435n
Locke, John, 96n
Lorenzen, Paul, 167, 200

Lycan, Bill, 490

Mac Lane, Saunders, 146, 147, 578
Mach, Ernst, 523
Maddy, Penelope, 288n
Martin, Donald, 440
McCarty, David Charles, 22, 394
McGee, Vann, 23
McLaughlin, Brian P., 613, 621
Meinong, Alexius, 579
Mellor, D. H., 592
Menne, Albert, 584
Meyerhof, Otto, 511
Mines, Ray, 233
Mirimanoff, Dimitry, 193
Montague, Richard, 141, 594
Moore, George E., 421, 432, 489, 577
Moore, Gregory, 3n, 350
Moss, Lawrence, 149n, 194
Muller, Frederik, 147n
Müller, Johannes, 529
Mycielski, Jan, 15, 19, 22, 153, 163f, 173, 177n, 549, 552n
Myhill, John, 195

Nelson, Edward, 181
Nelson, Leonard, 22, 509ff
Neurath, Otto, 496, 497
Niebergall, Karl-Georg, 19, 150

Palmgren, Erik, 232n
Parsons, Charles, 578
Parsons, Terence, 251
Pawlikowski, Janusz, 549
Peano, Giuseppe, 351, 379, 412, 508
Peckhaus, Volker, 2n, 5, 6n, 22
Peirce, Charles S., 453, 577
Planck, Max, 2
Plantinga, Alvin, 601
Plato, 484
Poincaré, Henri, 2, 5, 12, 191, 373, 533, 540, 577
Probst, Dieter, 18
Proclus, 102
Pudlák, Pavel, 158
Putnam, Hilary, 14, 551, 600

Quine, Willard V., 7, 14f, 136, 149, 273, 376, 495, 551, 580, 597, 602

Ramsey, Frank P., 14n, 349, 377, 379, 451, 464, 512, 594
Rao, Vidhyanath, 148n
Rathjen, Michael, 19, 232n
Reinhardt, Wiliam, 534
Resnik, Michael, 565
Richman, Fred, 231n, 237, 243n
Robinson, Abraham, 12
Rodríguez-Consuegra, Francisco, 8n, 21
Rouilhan, Philippe de, 21, 444n, 446n
Rüede, Christian, 119, 127
Rüstow, Alexander, 512f
Ruitenberg, Willem, 237
Russell, Bertrand, 1ff, 29, 49, 136ff, 191ff, 247f, 259f, 264f, 311, 349ff, 373ff, 401ff, 417ff, 435ff, 449ff, 481ff, 501, 507, 510f, 533, 543ff, 561ff, 567, 569ff, 577ff, 585ff, 594, 605, 611, 613, 615f, 619f, 639

Schlick, Moritz, 523
Scholz, Heinrich, 507
Schröder, Ernst, 507, 509, 577
Schroeder-Heister, Peter, 251
Schuster, Peter, 20
Schütte, Kurt, 13, 200
Schwichtenberg, Helmut, 20
Scott, Dana, 261n, 534
Sellars, Wilfrid, 599
Shapiro, Stewart, 551, 565
Shoemaker, Sidney, 607
Skolem, Thoralf, 452
Słupecki, Jerzy, 537
Sokal, Alan, 546

Solovay, Robert, 13
Specker, Ernst, 136, 150, 275
Steinhaus, Hugo, 534
Stout, George F., 424
St Paul, 349
Sturm, Holger, 23
Swoyer, Chris, 600

Tait, William, 159n, 569
Tarski, Alfred, 13, 404n, 537, 614
Thomas, Robert S. D., 23
Thomson, James, 168
Thomson, Judith J., 617n, 621
Tooley, Michael, 592, 607
Troelstra, Anne, 236n, 243
Twardowski, Kazimierz, 537

Unger, Peter, 611ff, 617n
Urmson, J. O., 496

van Bendegem, Jean Paul, 582n
van Dalen, Dirk, 243
van Inwagen, Peter, 613n
Vaught, Robert L., 141
Visser, Albert, 23
von Mises, Richard, 523
von Neumann, John, 352

Waismann, Friedrich, 171
Walton, Kendall, 581
Wang, Hao, 13, 200
Wasow, Thomas, 136n
Watson, John, 466
Wehmeier, Kai F., 20
Weierstrass, Karl, 375, 553
Weyl, Hermann, 13f, 192, 200, 451, 473, 512, 543
Wheeler, Samuel, 613n
Whitehead, Alfred N., 351, 375
Williams, C. J. F., 583
Williamson, Timothy, 618n
Wilson, Edward, 546
Wittgenstein, Ludwig, 423, 428, 430, 451, 464, 466, 494f, 523
Wood, Alan, 491
Woodin, W. Hugh, 16, 18, 101, 113, 187, 534
Wright, Crispin, 248n

Yablo, Stephen, 582
Young, William H., 509

Zalta, Edward, 248n
Zeno, 320, 353
Zermelo, Ernst, 2, 507ff

Subject Index

absolute undecidability, 525
absolutism, 23
AC, *see* axiom (scheme) of choice
acceptable ordinal, 301
acquaintance, 450
act
    of collection, 110
    of consciousness, 104
    of reflection, 110
actual infinite, 166, 549, 558
AD, *see* determinacy axiom
admissible set theory, 121
AFA, *see* anti-foundation axiom
algebra of logic, 507, 509
ambiguity, 135
    systematic, 135
    typical, 135
anti-foundation axiom (AFA), 19, 20, 149, 194
antinomy of infinite number, 354
approximate splitting principle, 229
arithmetic in ramified logics, 472
armageddon, 186
assumptions of infinity, 162
axiom (scheme)
    of infinity, 14n
    for propositions and truth, 261
    for the ε-operator, 537
    large cardinal, 31
    multiplicative, 438
    Newcomp, 50, 52
    of anti-foundation (AFA), 19, 20, 149, 194
    of choice (AC), 31, 438, 537
    of dependent choice (DC), 198, 227
    of determinacy (AD), 16, 33, 537
    of extensionality, 438, 443
    of inclusion subtlety, 52
    of infinity, 14n, 45, 438, 451
    of projective determinacy (PD), 16, 31, 32
    of reducibility, 14, 192, 418, 435, 437, 438, 449, 450, 469
    of reflection, 141
    of subtlety, 50
    of weak inclusion subtlety, 53
Axiom of Solvability, 518, 529
axiomatic method, 102
axiomatic program, 508
axiomatization, 502
    of logic, 509
    of set theory, 509

Baire property, 35
bar induction, 520
basic law V, 248
basic set theory, 121
Begriffsschrift, 247
Berry paradox, 378
β-conversion, 454
bisimulation, 203
Bolzano–Weierstrass theorem, 519
$\Box$ principle, 86
Bradley's paradox, 21, 419, 426, 429, 432
Bradley's regress, 423
British hegelianism, 353
British idealism, 489
Brouwer's intuitionism, 517
Burali-Forti paradox, 191, 349, 378

Cantor's Absolute, 96
Cantor's paradox, 1, 349, 350, 378, 502f, 506
Cantor's theorem, 20, 221, 249, 250, 260, 350

Cantorian set theory, 504
cardinal
  inaccessible, 96
  indescribable, 18, 51, 97
  large, 96
  Mahlo, 18, 51, 97
  measurable, 90, 97, 112
  strong, 97
  subtle, 18, 50f
  supercompact, 97
  superstrong, 89
  weakly compact, 51, 106
  Woodin, 33, 89, 90, 97, 99, 106
categorematical, 641
categorial grammar, 12
category of all categories, 145
category theory, 19, 23, 140, 145, 587
  axiomatic foundations, 146
  naive, 148
  structuralism based on, 565
circularity, 192
class
  proper, 306
  virtual, 49
class of all classes, 359
class symbol
  in Russell's substitutional theory, 387
  scope of, 439
class theory
  impredicative, 313
coinduction principle, 215
combinatory logic, 261
common knowledge, 193
complemented subset, 222
complex of entities, 419
comprehension, 248, 384
  elementary, 266
  naive, 285
  stratified, 273
  stratified explicit, 274
computer science, 11
concept, 248
conceptual realism, 587
conceptualist approach, 9
conditional
  material, 288
conditional platonism, 32
consciousness, 104
consistency proof, 251
  for arithmetic, 509
constituents, 430
constitution, 105
  of sets, 110
  of the concept of set, 110
constructible hierarchy, 75
constructible sets, 85, 99, 103
constructible universe, 13
  relativized to $\mathbb{R}$, 33
constructionalism, 449
constructive AFA set theory (CZFA), 198
constructive analysis, 227
constructive set theory, 195
constructive Zermelo–Fraenkel set theory (CZF), 196
constructivism, 12, 22, 412, 540
contemporary physics, 554
  fundamental assumption of, 554
context, 626
  in language, 626
  in logic and computer science, 628
context principle, 23
  Frege's, 625, 629
  truistic, 631
context_1, 626, 628
context_2, 628
context_3, 629
context/content duality, 628
context/text duality, 628
contextual definition, 393, 453
continuum, 322
  nonstandard, 526
continuum hypothesis, 18, 29, 44, 99
continuum problem, 93, 100, 113, 504

corecursion principle, 215
course-of-values, 248
cumulative hierarchy, 95
Curry paradox, 286
Curry property, 290, 298
cut-elimination, 462
CZF, *see* constructive Zermelo–Fraenkel set theory
CZFA, *see* constructive AFA set theory

DC, *see* dependent choice axiom
Dedekind infinite, 396
Dedekind real
  in constructive analysis, 241
definite description, 436, 453, 639
Δ^1_1-comprehension, 20
dependent choice axiom (DC), 198, 227
descriptive set theory, 16, 89
detachable subset, 222
determinacy axiom (AD), 16, 33, 537
Determinately (sentential operator), 614, 615
diagonal argument, 522
diagonalization, 20, 223, 527
  restricted, 278
disambiguation, 135
  conventions, 139, 143
discourse, 23, 637
discourse representation theory (DRT), 638
DRT, *see* discourse representation theory
dynamic predicate logic, 628

EFA, *see* exponential function arithmetic
effective topos, 224
Ehrenfeucht–Mostowski theorem, 150
elementary comprehension, 266
elementary extension, 254
Empiricist, 524, 527
endless regress, 424, 429
epistemic paradox, 464
ε-operator, 537 f
ersatz properties, 296
evidence, 102
excluded middle, 289
expansion
  sentential, 632
explicit mathematics, 148, 193
exponential function arithmetic (EFA), 50
extender model, 88
extensional simple type theory, 451
extensionality, 468
extensionality axiom, 438, 443
extra sentential, 631
extrapolation, 22, 549
  justification of, 553, 558
extrinsic evidence, 102

fallibilism, 22
fibered view on predicate logic, 628
fiction, logical, 401 f
fictionalism, 584
figural moment, 111
filling, 109, 112
Fin(T), 153, 163, 549, 557
Fin(ZF), 153
final coalgebra theorem, 217
fine structure theory, 86
finitarily justified system, 161
finite mathematics, 154, 165, 549, 557
  of the indefinitely large, 549
finite mathematics of the indefinitely large, 22
finitistic theory, 19, 153, 158 ff, 177
finitistically reducible, 153, 154 ff
fixed point approach to the semantic paradoxes, 288
fixed point lemma, 186
forcing, 31, 93, 101
form of a complex, 429
formal concept, 425
formalism, 534, 538, 543, 577
foundationalism, 22, 481, 491

founding, 107
four storey truth hierarchy, 319
free logic, 595
Frege Arithmetic, 11n
Frege structure, 261
Frege's context principle, 625, 629
Frege–Russell definition of number, 562
fulfillment, 107
functor category, 146
fundamental assumption of contemporary physics, 554
fundamental theorem of algebra, 236

Γ_0 (ordinal), 18, 200
Gentzen's Hauptsatz, 475
Gestalt-qualities, 111
Gödel fixed point, 182
Gödel's incompleteness theorems, 2, 159, 182
Gödel's completeness theorem, 37
Grelling's paradox, 510n, 512
Grundgesetze, 248

haecceities, 570
hermeneutic circle, 193
Hilbert circle, 507
Hilbert's paradox, 2, 22, 504ff, 510f
Hilbert's program, 159, 517ff, 551
Hume's Principle, 11n

Idealist, 524, 527
identity statement, 582
'if-then'ism, 564
Ignorabimus, 22, 523
impredicative, 254
impredicative class theory, 313
impredicative definitions, 191
inclusion subtlety, 52
incomplete symbol, 8, 385, 419, 436, 453
inconsistent multiplicities, 350
indescribable cardinal, 18, 51
individuals, 407f, 410
Infinitärcalcül, 522
infinitary mathematics, 551
infinite
  actual, 549, 558
  the actual, 166
  the potential, 154, 166
infinite game, 30, 98
  determined, 31, 98
infinitesimal number, 320
infinitesimals, 522, 525, 527
infinity axiom, 45, 438, 451
inscrutability of reference, 612, 621
inseparability, 268
insolubilia, 376
intensional interpretation, 435
intensional logic, 12, 464, 594
intentional paradox, 464
intentionality, 104
interactive prover Coq, 236
intermediate theory (Russell), 21, 401ff
  formal language of, 408f
intermediate value theorem, 234
intra-sentential, 631
intrinsic evidence, 102
intuition, 94, 542
intuitionism, 22, 517, 530, 540, 543
intuitive act, 107
iterative concept of set, 95, 111

justification of extrapolation, 553, 558

knowledge by acquaintance, 483
knowledge by description, 483
König/Dixon paradox, 378
KP, *see* Kripke–Platek set theory
Kripke's theory of truth, 21, 193, 288, 317n, 595
Kripke–Feferman self-referential truth, 278
Kripke–Platek set theory (KP), 18, 120, 199

L, 112
lambda abstraction, 261

large cardinal, 96
large cardinal axioms, 31
large cardinal hierarchy, 100
large ordinal, 315
largest fixed point, 209
law of excluded middle, 289
Lebesgue measurability, 16, 30
Liar paradox, 2, 191, 317, 379, 505, 512
limit ordinal, 315
linguistic paradox, 349
litorial, 315
locally finite theory Fin(T), 163
logical atomism, 8, 21, 417, 485
logical paradox, 376
logical product, 259
logical syntax, 7
logicism, 14, 543
LR, *see* reflexion logic

Mahlo cardinal, 18, 51, 97
many-valued logic, 614
Martin-Löf type theory, 12, 196, 202, 629
material conditional, 288
  paraconsistent approach, 288
mathematical induction, 22
  Russell's proof of, 452, 472
mathematical object, 23, 105, 584
mathematical practice, 579
mathematics
  explicit, 148, 193
  finite, 154, 165, 549, 557
  infinitary, 551
  reverse, 15
measurable cardinal, 90
measurement, 552, 555 ff
  scientific, 22
mental construction, 535
mental mechanism, 536
mereology, 617
metapredicative Mahlo, 19, 119, 122
metatheory, 7
modal operators, 168
modal paradox, 260
monism, 419
Morse–Kelley set theory, 308
move in the language-game, 630
multi-valued semantics, 21, 292
multiple relation theory of judgement, 420, 431, 453
multiplicative axiom, 438
Mycielski-translation
  into Fin(ZF), 173

naive category theory, 148
naive comprehension, 285
naive theory of properties, 285
natural language philosophy, 495
natural ontology, 592, 606
natural property, 309
naturalistic conception of properties, 607
naturalized epistemology, 495
NBG, *see* Neumann–Bernays–Gödel set theory
Nelson circle, 509 f
neo-Hegelian, 4
Neue Fries'sche Schule, 510
Neumann–Bernays–Gödel set theory (NBG), 12, 86, 308
neutral monism, 375
New Foundations (NF), 19, 20, 136, 273, 535
  with urelements (NFU), 149, 260
  subsystems, 275 ff
Newcomp (comprehension axiom scheme), 50, 52
NF, *see* New Foundations
NFU, *see* New Foundations with urelements
no-classes theory, 9, 192, 265, 435
noema, 105, 109, 114
nominalism, 15, 597
non-standard analysis, 12
non-well-founded sets, 193

nonstandard continuum, 526
number 0
  definition of, 441

object sense, 105, 106
objective, 112
objective world, 554
objectivity, 104
objectual quantification, 374
Ockham's Razor, 8
Ω Conjecture, 37
Ω-logic, 29, 34, 101, 113
Ω-proof, 36
Ω-soundness, 36
ontological atom, 484
ontological commitment, 597
ontological development, 466
ontological eliminativism, 374
ontology, 21
open (truth value), 317
order (vs. type), 405
order completeness, 232, 234
order indiscernibles, 556
ordinal
  acceptable, 301
  Γ_0, 18, 200
  large, 315
  limit, 315
  proof-theoretic, 119
  subtle, 52
  successor, 315
  transdefinite, 315
ordinal analysis, 18

paradox
  Berry, 378
  Bradley, 21, 419, 426, 429, 432
  Burali-Forti (greatest ordinal), 191, 349, 378
  Cantor (greatest cardinal), 1, 349, 350, 378, 502f, 506
  Curry, 286
  epistemic, 464
  Grelling, 510n, 512
  Hilbert, 2, 22, 504ff, 510f
  intentional, 464
  König/Dixon, 378
  Liar, 2, 191, 317, 379, 505, 512
  linguistic, 349
  logical, 376
  modal, 260
  of propositions, 20, 377
  po/ao, 377, 388
  recurrence, 2
  Richard, 378
  Russell, 1, 29, 49, 137, 191, 247, 265, 285, 349, 350, 501f, 505, 511, 512, 593, 600, 605
    in a constructive context, 221
    predicational variant, 364
  selfreferential, 349
  semantic, 317, 349, 376, 464
    fixed point approach, 288
  set-theoretic, 464
  Skolem, 534
  Zermelo, 501, 505, 507, 510
paradox of propositions, 401
patterns
  mathematical structures as, 565
PD, *see* projective determinacy axiom
perception, 106
perspective variations, 106
phenomenology, 104, 114
physical quantity, 554, 556
pigeon hole principle, 476
platonism, 21, 94, 311, 313, 449, 538, 551, 577
  conditional, 32
  Gödel's, 12
po/ao paradox, 377, 388
possible world semantics, 601
postulationism, 586
potential infinite, 154, 166
PRA, *see* primitive recursive arithmetic
pre-utterance, 628
predicable, 1, 364

predicate logic, 628
dynamic, 628
fibered view, 628
predicative, 252
predicative analysis, 473
predicative class hierarchy, 319
predicativism, 12, 200
system W, 14
predicativity, 402
primitive recursive arithmetic (PRA), 19, 155
principle of disjunction (of types), 404
principles for propositions and truth, 261
problem of the many, 612
projective determinacy axiom (PD), 16, 31, 32, 98, 102, 112
projective sets, 29, 98
measure problem, 44
proof-theoretic ordinal, 119
proof-theoretical reducibility, 154
proper class, 306
property
Baire, 35
natural, 309
naturalistic conception, 607
semantical conception, 591
semantically-conceived, 309
property theories, 595
proposition, 7, 419, 450
elementary, 405ff
elimination of, 406ff
general, 404f
hierarchy of, 404f, 410ff
paradox of, 20
propositional function, 8, 418, 450
propositional object, 261

quantification, 9
objectual, 374
substitutional, 450, 459
Quine–Putnam indispensability, 14, 551

ramified analysis, 272
ramified second-order logic, 457
ramified theory of types, 10, 450
ramified third-order logic, 457
ramified types, 457
rationalism, 22, 535, 538f
real number, 324
in constructive analysis, 227
transdefinite, 345
transfinite, 329, 336
recurrence paradox, 2
recursively saturated, 254
reducibility
finitistic, 154
proof-theoretical, 154
reduction
of material objects
to sense-data, 483
reductionism, 14
reflected truth, 318
reflection principle, 97, 111
reflective universes, 141
reflexion logic (LR), 21, 317
regressive method, 15, 491
relations, 418
resource consciousness, 15
restricted diagonalization, 278
reverse mathematics, 15
Richard paradox, 378
Robinson's arithmetic, 253
Russell's eliminativism, 375
Russell's method of analysis, 487
Russell's paradox, 1, 29, 49, 137, 191, 221, 247, 265, 285, 349, 350, 501f, 505, 511, 512, 534, 593, 600, 605
in a constructive context, 221
predicational variant, 1, 364
Russell's substitutional theory, 373ff
axiom schemata for the 1905 version, 380
axiom schemata for the 1906 version, 390

Russell's theory of knowledge, 22, 482
Russell–Zermelo paradox, *see* Russell's paradox
Russell-the-fallibilist, 481
Russell-the-foundationalist, 481

satisfaction, 306
schema V, 254
scope
  of class symbols, 439
second-order logic, 248
second-order number theory, 30, 98
self-application, 591
self-instantiation, 591, 592, 606
self-predication, 1, 11
self-referential imperative, 312, 324
self-referential paradox, 349
self-referential truth
  Kripke–Feferman, 278
semantic paradox, 193, 317, 349, 376, 464
semantic pretence, 582
semantical conception of properties, 591
semantically unfounded sentence, 317
semantics, 23, 592, 601
  3-valued, 292
  multi-valued, 21, 292
  realist, 584
  substitutional, 469
sense-data, 483ff
sentence, 640
  definition of, 634
sentential expansion, 632
sequential completeness, 231, 234
set
  constructible, 85, 99, 103
  non-well-founded, 193
  projective, 29, 44, 98
  virtual, 49, 61
set continuous class operator, 208
set theory
  New Foundations (NF), 535
  admissible, 121
  axiomatization of, 509
  basic, 121
  Cantorian, 504
  constructive, 195
    CZFA, 198
    Zermelo–Fraenkel (CZF), 196
  descriptive, 16, 89
  Kripke–Platek (KP), 18, 120, 199
    KPi^0, 18, 120
  Morse–Kelley, 308
  Neumann–Bernays–Gödel (NBG), 12, 86, 308
  New Foundations (NF), 19, 20, 136, 273
    subsystems, 275ff
    with urelements (NFU), 149, 260
  with universes, 141
  Zermelo Z, 11
  Zermelo–Fraenkel (ZF), 11, 13, 33, 141, 596
    with choice (ZFC), 34
    with reflective universes, 141
    with the Axiom of Choice (ZFC), 141
set-theoretic paradox, 464
sharp operation, 18
Σ operation, 119, 122
signitive act, 107
simple theory of types, 11, 137, 438, 606
simultaneous dual substitution, 385
single-candidate thesis, 617ff
Skolem paradox, 534
solution lemma, 207, 212
stratification, 273
stratified comprehension, 273
stratified explicit comprehension, 274
stratified truth, 273
stratified T-schema, 280
streams, 213
structural realism, 374
structuralism, 23, 587

ante rem, 565
Dedekind's, 23, 562, 569
eliminative, 565
modal, 565, 587
set-theoretic, 565
subcountable, 223
subject and predicate, 426, 430
subjective, 112
subjectivity, 104
substitution, 7, 376
simultaneous dual, 385
substitutional interpretation, 452
substitutional quantification, 450, 459
substitutional semantics, 469
substitutional theory (Russell), 21, 373 ff, 402, 404
subtle cardinal, 18, 50 f
subtle ordinal, 52
successor ordinal, 315
superstrong cardinal, 89
symbolic explicitness, 6
syncategorematical, 641

T-schema
stratified, 280
tertium non datur, 519
theory of intuition, 542
theory of judgment, 406 ff, 418, 426
theory of properties, 23
naive, 285
theory of truth, 13, 20, 595
Kripke, 21, 193, 288, 317 n, 595
theory of types, 10
Martin-Löf, 12, 196, 202, 629
of attributes
axiom schemata for, 384
ramified, 10, 402, 450
simple, 11, 137, 438, 606
extensional, 451
third truth value, 317
thought-objects, 533
3-valued semantics, 292
topoi, 567
transcendence, 109
transcendental, 113
transcendental predicate, 594
transdefinite ordinal, 315
transdefinite real numbers, 345
transdefinite well-ordering, 314
transfinite real numbers, 329, 336
truistic context principle, 631
truth, 541
hierarchy, 319
reflected, 318
self-referential, 278
stratified, 273
theory of, 595
unreflected, 318
truth value gaps, 614
twofold nature of relations, 426
type, 403 f
of level n, 467
ramified, 457
simple, 466
type free language, 595
type symbol, 383
typed λ-calculus, 456
types
doctrine of, 401
ramified theory of, 402
typical ambiguity, 19, 135

uncountable, 223
unfounded sentence, 317
unit of meaning, 23, 630, 637
universalism, 403 f
universally Baire, 35
universals, 424
unreflected truth, 318

V, 95, 103, 113
vagueness, 23, 613
value-range, 248
Veblen function, 19 n, 201
ternary, 129
$V = L$, 61, 86

vicious circle principle, 10, 12, 192, 374, 377, 402ff
vicious regress, 192
virtual class, 49
virtual set, 49, 61
von Neumann hierarchy, 85

Wadge hierarchy, 36
weak counterexamples, 524
weak inclusion subtlety, 53
weakly compact cardinal, 51
well-ordering
    transdefinite, 314
Wiener–Kuratowski pair, 453
Woodin cardinal, 33, 89, 90, 97
word
    definition of, 634

Zeno's paradox, 320ff
Zermelo set theory Z, 11
Zermelo's paradox, 501, 505, 507, 510
Zermelo–Fraenkel set theory (ZF), 11, 13, 33, 596
    with choice (ZFC), 34, 93, 110
$0^{\#}$ exists (large cardinal axiom), 88, 90
ZF, *see* Zermelo–Fraenkel set theory
ZFC, *see* Zermelo–Fraenkel set theory with choice